MALIGNANCY-RISK STRATIFICATION OF THYROID NODULES: DIAGNOSTIC UTILITY OF THE ULTRASOUND COMPUTER-AIDED DIAGNOSIS AND ADVANCED ULTRASOUND IMAGING TECHNIQUES

NONHLANHLA CHAMBARA

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Health Technology and Informatics

Malignancy-Risk Stratification of Thyroid Nodules: Diagnostic Utility of the Ultrasound Computer-Aided Diagnosis and Advanced Ultrasound Imaging Techniques

NONHLANHLA CHAMBARA

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

April 2022

Certificate Of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

Nonhlanhla Chambara (Name of Student)

Abstract

Thyroid cancer is the most prevalent endocrine cancer which typically manifests as thyroid nodules that are primarily diagnosed using ultrasound and fine-needle aspiration cytology (FNAC). The surge in thyroid cancer incidence and the overdiagnosis of indolent thyroid nodules in recent years is attributed to the increased use of advanced ultrasound technology and FNAC. This thesis evaluated the diagnostic utility of computer-aided diagnosis (CAD), AngioPLUS microvascularity assessment and shear wave elastography (SWE) in thyroid cancer diagnosis.

Study 1 validated the diagnostic performance of CAD software based on 6 TIRADS using 205 thyroid nodules at the default setting and 3 adjusted settings. The results confirmed that the default setting was ideal for achieving maximum sensitivity. The EU TIRADS demonstrated the highest sensitivity (82.7%). Comparative diagnostic performance analysis was done between the CAD and computer-assisted subjective interpretation using 4 analogous TIRADS and 162 nodules. The results showed comparable sensitivity between the two methods, however, CAD had lower specificity for all TIRADS. EU TIRADS and KSThR TIRADS differentiated all nodules with comparable diagnostic performance, however, EU TIRADS had slightly higher specificity.

Study 2 evaluated the diagnostic value of AngioPLUS coupled with conventional power (APDI) and colour Doppler (ACDI) ultrasound imaging techniques in combination with EU TIRADS. A total of 94 thyroid nodules that included 40 cytologically-equivocal nodules were analysed using a quantitative ratio vascularity index (RVI) at different offsets and qualitative visual regional grading. RVI combined with EU TIRADS resulted in high specificity but poor sensitivity at all offsets and Doppler ultrasound modes. The combination of qualitative APDI

vascularity grading and EU TIRADS had higher specificity than EU TIRADS alone and slightly lower sensitivity in differentiating all nodules (SEN: 76.7% vs 83.3%, $p > 0.05$, SPEC: 84.4% vs 50%, $p < 0.05$;). However, it stratified cytologically-equivocal nodules with significantly higher overall diagnostic performance than EU TIRADS alone (SEN: 88.9%; SPEC: 77.4% vs 38.7%, $p < 0.05$, and AUROC: 0.89 vs 0.62, $p < 0.05$).

Study 3 evaluated the diagnostic value of SWE in combination with EU TIRADS and included 126 thyroid nodules. Using different sub-categories of the nodules, diagnostic performance analyses were performed. Results showed that no SWE index was superior to EU TIRADS in differentiating all nodules. However, for nodules between 1cm and 2 cm at an optimal cut-off of 8.7 kPa and those > 2cm at an optimal cut-off of 10.7 kPa, the SWE SD index combined with EU TIRADS demonstrated higher specificity and insignificantly lower sensitivity than EU TIRADS alone (SEN:72.2% vs 88.9%, $p > 0.05$; SPEC:76.5% vs 55.9%, $p < 0.05$; and, SEN: 71.4% vs 85.7%, $p > 0.05$; SPEC: 95.8% vs 62.5%, $p < 0.05$, respectively).

In conclusion, CAD has a complementary diagnosis role in thyroid cancer screening, but it is not superior to subjective interpretation. EU TIRADS can rule in thyroid malignancy using either CAD or subjective interpretation. APDI combined with EU TIRADS is highly diagnostically efficient in differentiating cytologically-equivocal thyroid nodules. SWE combined with EU TIRADS has the potential for improving the diagnosis of nodules ≥ 1cm.

Presentations and Publications

Conference Presentations

Chambara N, Ying M, Liu S, Lo X. Computer-assisted subjective analysis of thyroid nodules may limit nodule non-specification than computer-aided diagnosis. 19 Asian Oceanian Congress of Radiology (AOCR)- From Pixel to Clarity, 1st – 4th July 2021, Kuala Lumpur, Malaysia. (Abstract, Oral Presentation) **[Selected as one of the finalists for the Top Free Paper Oral Presentation]**

Publications

1. **Chambara N**, Ying M. The Diagnostic Efficiency of Ultrasound Computer-Aided Diagnosis in Differentiating Thyroid Nodules: A Systematic Review and Narrative Synthesis. *Cancers (Basel)*. 2019;11(11):1759. [2020 Journal Impact Factor = 6.639; Best ranking = 51/242]
2. **Chambara N**, Liu SYW, Lo X, Ying M. Diagnostic performance evaluation of different TI-RADS using ultrasound computer-aided diagnosis of thyroid nodules: An experience with adjusted settings. *PLoS One*. 2021;16(1):e0245617. [2020 Journal Impact Factor = 3.24; Best ranking = 26/72]
3. **Chambara N**, Liu SYW, Lo X, Ying M. Comparative Analysis of Computer-Aided Diagnosis and Computer-Assisted Subjective Assessment in Thyroid Ultrasound. *Life (Basel)*. 2021;11(11):1148. [2020 Journal Impact Factor = 3.817; Best ranking = 27/93]

4. **Chambara N**, Lo X, Liu SYW, Ying M. Diagnostic value of AngioPLUS microvascular imaging in thyroid nodule diagnosis using quantitative and qualitative vascularity grading. *Biomedicines* 2022, 10(7), 1554. [2021 Journal Impact Factor = 4.757; Best ranking = 30/138]
5. **Chambara N**, Lo X, Liu SYW, Ying M. Combined shear wave elastography and EU TIRADS in differentiating malignant and benign thyroid nodules. (To be submitted to *Cancers*)

Acknowledgements

I greatly appreciate Professor Michael YING, my supervisor, for his professional guidance and mentorship throughout my PhD studies. I am indebted to him for equipping me with skills and knowledge by sharing his expertise in diagnostic ultrasound imaging techniques and research. His valuable advice, support and continual motivation enabled me to forge ahead and complete this thesis despite various challenges that I encountered over the past 4 years.

My gratitude also goes to Drs. Shirley Y.W. Liu and Xina Lo for their valuable contribution through their expertise in thyroid surgery and assistance with data collection and patient recruitment at the Prince of Wales Hospital (PWH) Department of Surgery. I also extend my gratitude to the other staff at the PWH Department of Surgery and Ms. Emma Chow for assisting with recruiting and scheduling the patients on a week-to-week basis while the project was ongoing.

I would also like to thank my colleagues from the Department of Health Technology and Informatics. Notably, I appreciate Dr. Yuanxi Li, Mr. Howard Lin and Mr. Bosco Chou for their kind help in communicating with the patients during the data collection process; Ms. Adenike Adesanya, Ms. Regina Kwarteng and Dr. Li Jia for their encouragement, advice, and support throughout this academic journey.

I appreciate The Hong Kong Polytechnic University and Department of Health Technology and Informatics for the provision of the research studentship grant, resources and personnel that enabled me to undertake my PhD studies and expand my knowledge and skillset at this reputable institution.

Last but not least, I am deeply grateful to my mom, the rest of my family and friends for their prayers, love, encouragement, and tremendous support during this academic journey.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AACE	American Association of Clinical Endocrinologists
ACE	American College of Endocrinology
ACFI	AngioPLUS + Colour Flow Imaging
ACR	American College of Radiology
AME	Associazione Medici Endocrinologi
AngioPLUS	Angio Planewave Ultrasensitive
APDI	AngioPLUS + Power Doppler Imaging
ATA	American Thyroid Association
ATC	Anaplastic Thyroid Cancer
AUROC	Areas Under the Receiver Operating Characteristic Curve
AUS	Atypia of Undetermined Significance
BTA	British Thyroid Association
CAD	Computer-aided diagnosis
CDUS	Colour Doppler Ultrasound
CEA	Carcino-embryonic antigen
CEUS	Contrast enhanced ultrasound
CFI	Colour Flow Imaging
CI	Confidence Interval
CT	Computed Tomography
DA	Diagnostic Accuracy
DOR	Diagnostic Odds Ratio
DTC	Differentiated Thyroid Carcinoma
EU	European Thyroid Association
^{18}F-FDG PET	Fluorodeoxyglucose Positron Emission Tomography
FLUS	Follicular Lesion of Undetermined Significance
FN	Follicular Neoplasm

FNAC	Fine Needle Aspiration Cytology
FTC	Follicular Thyroid Cancer
HCC	Hürthle Cell Cancer
^{123}I	iodine-123
^{131}I	Iodine-131
kPa	kiloPascals
KSThR	Korean Society of Thyroid Radiology
KTA	Korean Thyroid Association
MEN	Multiple Endocrine Neoplasia
MRI	Magnetic Resonance Imaging
MTC	Medullary Thyroid Cancer
NaIS	Sodium-Iodine Symporter
NIFTP	Non-Invasive Follicular Thyroid Neoplasm with Papillary-like Nuclear Features
NLR	Negative Likelihood Ratio
NPV	Negative Predictive Value
PDI	Power Doppler Imaging
PDUS	Power Doppler Ultrasound
PET	Positron Emission Tomography
PLR	Positive Likelihood Ratio
PPV	Positive Predictive Value
PRC	Precision Recall Curve
PTC	Papillary Thyroid Cancer
PTMC	Papillary Thyroid Microcarcinoma
RAI	Radio-active Iodine
ROC	Receiver Operating Characteristic curve
ROI	Region of Interest
RSS	Risk-Stratification System
RTE	Real Time Elastography

RVI	Ratio Vascularity Index
SE	Strain Elastography
SEN	Sensitivity
SFN	Suspicious for a Follicular Neoplasm
SMI	Superb Microvascular Imaging
SPEC	Specificity
SUV	Standard Uptake Values
SWE	Shear Wave Elastography
^{99m}Tc	Technetium
$^{99m}TcO_4$	Technetium Pertechnetate
T_3	Triiodothyronine
T_4	Thyroxine
Tg	Thyroglobulin
TIRADS	Thyroid Imaging Reporting and Data System
TRH	Thyrotropin-Releasing Hormone
TSH	Thyroid-Stimulating Hormone
VI	Vascularity Indices

Chapter 1

Introduction

Thyroid cancer is the most common endocrine carcinoma globally, making up 3% of all cancer malignancies [1]. The incidence of thyroid cancer has increased over the years with recent studies in China showing a moderate increase in morbidity among women with increasing age [2]. In Hong Kong, it is the second most common cancer affecting young women [3]. Thyroid cancer commonly presents as symptomatic or asymptomatic thyroid nodules of which only about 15% tend to be malignant [4]. The pre-surgical diagnosis of thyroid nodules is usually through diagnostic ultrasound imaging. Depending on the sonographic features and the nodule size, fine needle aspiration cytology (FNAC) may also be conducted for diagnosis. The rise in thyroid cancer incidence has been attributed to the advancement and increased sensitivity of diagnostic ultrasound imaging modalities coupled with the increased use of FNAC. A challenge that has arisen with the increase in thyroid cancer incidence is the overdiagnosis of indolent or low-risk thyroid neoplasms without a significant influence on the mortality rates [4, 5].

Although FNAC is considered the minimally-invasive pre-operative reference standard for thyroid nodule diagnosis, it has the drawback of about 30% equivocal results for which only up to 30% of the nodules are malignant [6-8]. Equivocal cytology specimens currently can only be conclusively diagnosed by repeat FNAC or post-thyroidectomy histopathology. This implies that potentially, patients with benign nodules may be subjected to unnecessary thyroidectomy for a conclusive diagnosis. This in turn has surgical complication risks as well as cost and quality of life implications that come with a lifetime of thyroid hormone replacement therapy. Therefore, there is a need for less invasive thyroid nodule diagnosis

approaches that are complementary and diagnostically-efficient for more conservative treatment of otherwise benign nodules.

Ultrasound is the primary diagnostic imaging modality in differentiating benign and malignant thyroid nodules. The ultrasound diagnosis of thyroid nodules depends on their morphological appearance on a sonogram. Typical grey scale ultrasound features that suggest thyroid malignancy include marked hypoechogenicity, taller than wide shape, irregular margins and microcalcifications, while for benign disease they include spongiform appearance, peripheral halo, macrocalcifications and regular shape/margins [9-12]. However, as no grey scale ultrasound feature accurately predicts malignancy with great precision singularly, a combination of suspicious features have been suggested to increase sensitivity but this often results in lower specificity [13-15]. Thyroid Imaging Reporting and Data System (TIRADS) guidelines based on the stratification of the risk of malignancy using several predictive sonographic features have emerged in different countries. However, different guidelines vary in malignancy-risk criteria and estimation even among shared common predictive features thereby resulting in the lack of consensus on the ideal TIRADS to use to date. Due to the inter and intra-observer variances that arise from the subjective interpretation of ultrasound images, thyroid ultrasound computer-aided diagnosis (CAD) techniques that are purported to be more objective have emerged. Although the globally-approved commercial thyroid ultrasound CAD (AmCAD-UT, AmCad Biomed, Taipei, Taiwan) incorporates multiple TIRADS, the comparison of the diagnostic performance of the different TIRADS in diagnosing the same thyroid nodules to ascertain the optimal TIRADS is lacking in previous studies. A diagnostic performance validation of the CAD software and a comparison between the diagnostic efficacy of the CAD software and computer-assisted subjective interpreters for the same multiple TIRADS is presented in study one in Chapter 3.

Thyroid ultrasound imaging is not limited to grey scale ultrasound assessment, but also involves the assessment of thyroid nodule vascularity. Angiogenesis manifests in malignant tumours as the highly metabolic carcinogenic cells invade, grow and create new chaotic and torturous blood vessels for survival in areas deficient in blood vessels [16, 17]. This results in proliferation and increased vascularity in tumour areas. Contentions persist regarding the role of vascularity assessment in determining thyroid nodule malignancy. However, some previous studies have indicated that increased chaotic central vascularity raises suspicion of malignancy, whereas peripheral vascularity is suggestive of the benign disease process [18-20]. Colour and power Doppler ultrasound are commonly used to assess the vascularity in thyroid nodules. However, these conventional modes are poor at detecting low blood flow in microvessels. The advent of advances in ultrasound technology has resulted in the emergence of novel microvascularity assessment techniques. Angio Planewave Utrasensitive (AngioPLUS) imaging is one such ultrafast Doppler imaging technique with high sensitivity and image resolution based on spatiotemporal filtering for the increased detection of low blood flow in microvasculature [21]. AngioPLUS has shown effectiveness in differentiating parathyroid lesions from other lesions, however, reporting of its utility in thyroid imaging is scant [22]. Owing to its increased sensitivity and image resolution in detecting microvascular flow, its diagnostic performance can be speculated to be superior to conventional Doppler ultrasound.

In thyroid ultrasound, the commonly used methods of vascularity assessment are qualitative grading based on various pattern analyses. The value of Doppler ultrasound in thyroid nodule assessment could be influenced by the Doppler ultrasound techniques used, and whether the vascularity grading approach is qualitative or quantitative assessment. In a previous study, our group demonstrated the utility of a customised regional segmentation algorithm based on offsetting in the quantitative analysis of thyroid nodule vascularity when combined with grey scale ultrasound [23]. There is a lack of investigation of the diagnostic efficacy of combination

approaches involving TIRADS and Doppler ultrasound assessment in thyroid nodule diagnosis. Furthermore, the evaluation of the diagnostic value of such approaches in improving the diagnosis of thyroid nodules with equivocal cytology is lacking. Chapter 4 presents study two in which a ratio vascularity index (RVI) based on thyroid regional nodule vascularity segmentation at different offsets was used in quantitative vascularity analysis for the first time. Based on the RVI and qualitative vascularity assessment, the diagnostic performances of AngioPLUS coupled with colour and power Doppler ultrasound were determined and compared to those of sole colour and power Doppler ultrasound. The study also investigated the diagnostic performances of different Doppler ultrasound modes in combination with EU TIRADS for the assessment of all nodules and nodules with equivocal cytology.

The carcinogenesis process can manifest with stiffness changes along with other morphological features and vascularity changes. Tissue elasticity is a physical parameter that is mostly affected by the pathophysiological process [24]. Most of the malignant lesions tend to have increased stiffness while benign lesions are more elastic. In papillary thyroid cancer, the most common type of thyroid cancer, the increased stiffness is attributed to fibroblastic reactions and presence of psammoma bodies and increased cellular compaction caused by the infiltrative nature of cancer [25]. Elastography is the ultrasound imaging modality used in the assessment of tissue elasticity changes based on determining the tissue displacement changes in response to extrinsic or intrinsic pressure. Shear wave elastography (SWE) is one of the recent advances in elasticity imaging that is considered operator-independent and more accurate in assessing tissue elasticity. SWE uses acoustic radiation impulses whereby the velocity of the shear waves in tissue is converted to Young's modulus to determine tissue elasticity and generate tissue stiffness values in absolute quantitative measurements in kiloPascals (kPa) [26].

Although several authors have reported that in combination with grey scale ultrasound assessment, SWE can help triage thyroid nodules for biopsy, there are contrasting views

regarding its diagnostic accuracy in thyroid nodule diagnosis [27, 28]. The major drawbacks in its clinical utility have been the lack of universal standards regarding the optimal scanning planes, optimal SWE indices, and optimal cut-off points of stiffness values for use in determining thyroid malignancy [29-31]. Furthermore, there is currently no consensus regarding the influence of nodule size on SWE diagnostic performance outcomes. Elastography has been purported to improve the specificity for pre-surgical triaging of patients with cytologically equivocal results [32]. However, the investigation of the diagnostic efficacy of SWE in diagnosing nodules with equivocal cytology is limited and the current few studies have shown conflicting results [33, 34]. Multimodal studies involving SWE have been suggested to have the potential in improving the diagnostic accuracy of grey scale ultrasound alone, particularly for nodules with equivocal cytology, however, very few studies have evaluated this [35]. The triaging of patients with thyroid nodules for appropriate treatment depend on optimised thyroid ultrasound diagnosis approaches which in turn may limit unnecessary treatment procedures. In Chapter 5, the diagnostic efficacy of SWE in combination with EU TIRADS was investigated using a pre-determined optimal scan plane with the corresponding SWE indices based on sub-categories of nodule size, equivocal cytology, and vascularity status. The diagnostic value of a multimodal approach in thyroid nodules with equivocal cytology was determined using the combination of EU TIRADS with AngioPLUS coupled with power Doppler and SWE.

Chapter 2

Literature Review

2.1 Brief overview of the anatomy and physiology of the thyroid gland

The thyroid gland is an endocrinal gland that is responsible for the production of the thyroid hormones for the regulation of metabolism in the body. Its secondary role is the production of calcitonin for calcium metabolism. It is located at the level of the third and fourth cervical vertebrae anterior to the trachea. This proximity to the trachea can often result in posterior nodules being missed during routine clinical examinations [36]. The thyroid gland consists of two oval-shaped lateral lobes connected by an isthmus centrally, and the thyroid lobes are usually symmetrical, about 10 to 20 grams and with an average size of 5cm by 5cm in adults (Figure *2*:*1*) [37]. Occasionally a pyramidal lobe can be found in some individuals, most often children. The thyroid gland is medial to the common carotid artery and internal jugular vein on both sides and the oesophagus is found postero-lateral to it on the left side. The thyroid gland's main arterial supply is from the superior thyroid artery and inferior thyroid artery. The venous drainage is by the superior and middle thyroid veins which drain into the internal jugular vein and inferior veins which then anastomose and drain into the brachiocephalic vein. The lymphatic drainage is to the central and bilateral cervical lymph nodes [38].

Follicles form the structural and functional unit of the gland. The follicular cells make up the surface of the epithelium and are responsible for thyroid hormone synthesis while the parafollicular C- cells are responsible for calcitonin production [38, 39].

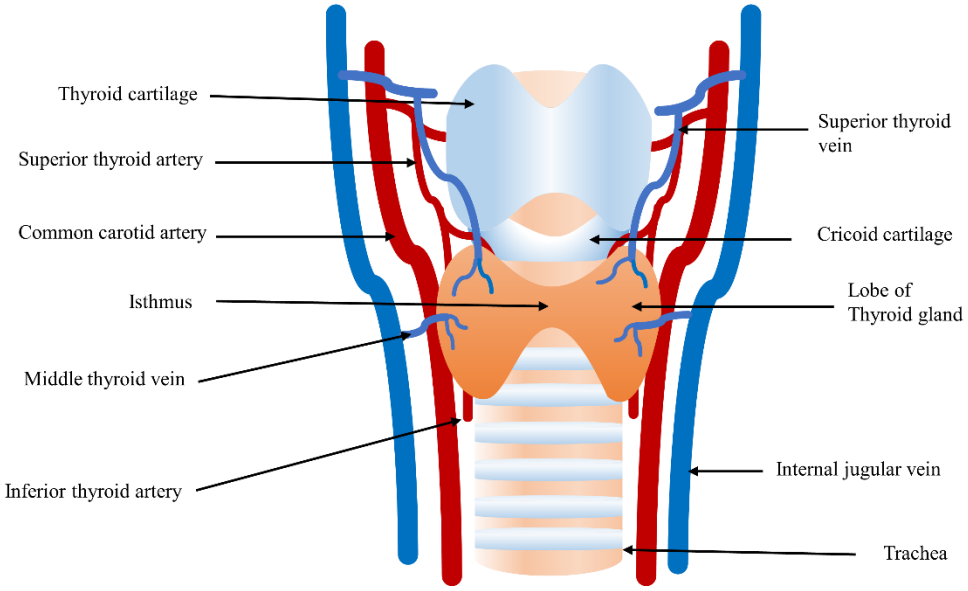


Figure 2:1 A schematic representation of the thyroid gland and surrounding structures and blood supply

In each follicle, the hormones are stored as colloid. Since the thyroid hormone is iodinated, the thyroid gland concentrates and organifies iodine from circulation through the action of a sodium-iodine symporter (NaIS) that is stimulated by thyrotropin. A normal thyroid cell should be able to trap iodine from blood plasma against a concentration gradient; this property influences radioactive iodine (RAI) use to destroy thyroid cancer cells [40]. The thyroid hormone binds to nuclear receptors in target cells and influences the expression of regulatory genes that regulate basal metabolic rate, protein synthesis as well as fat and carbohydrate metabolism [38]. Thyroid hormone output consists of about 90% (75-100ug/day) thyroxine (T_4) and about 10% (6ug/day) triiodothyronine (T_3) [36, 41]. Thyroid hormones (T_4 and T_3) are therefore regulated by thyroid-stimulating hormone (TSH) which in turn is regulated by the thyrotropin-releasing hormone (TRH) from the hypothalamus as well as feedback inhibition by the thyroid hormones. TSH and free thyroxine T_4 (fT_4) have an inversely log-linear relationship such that a minor change in fT_4 is accompanied by a large reciprocal change in TSH [42]. This relationship is critical in the diagnosis and treatment of thyroid abnormalities.

2.2 Thyroid cancer

2.2.1 Thyroid cancer epidemiology

Thyroid cancer is the most common endocrine malignancy which accounts for about 3% of all cancers [1]. The most common presenting feature of thyroid cancer is thyroid nodules which can be present in 1% of men and 5% of women [38, 43]. Thyroid nodules can be found incidentally during medical imaging examinations; however, over 90% of thyroid nodules are usually benign [43-45]. Up to 50% of adults above 50 years old can harbour one or multiple non-clinically relevant sub-centimetre thyroid nodules, with most being discovered incidentally on autopsies and resulting in subsequently increased thyroid microcarcinoma rates [46, 47].

Thyroid cancers are more prevalent in women and are the fifth most common malignancy in women accounting for about 5% of all female cancer diagnosis worldwide [1, 48, 49]. Based on recent studies, there is a gradual increase in thyroid cancer morbidity in China, particularly among female patients who are almost 4 times affected by thyroid cancer than male patients [2, 50]. In Hong Kong, thyroid cancer is the 2nd most common cancer in women after breast cancer in the age group of 20 to 44 years [3]. The high incidence of thyroid cancer in women suggests that female hormones may have a regulatory role in thyroid carcinogenesis [39].

2.2.2 Common histological variants of thyroid cancer

About 90% of all thyroid cancers are of follicular cell origin while 4% originate from C-cells [39, 51]. Papillary thyroid cancer (PTC) which accounts for over 80% of all thyroid cancers, follicular thyroid cancer (FTC, up to 11%), Hürthle cell cancer (HCC) and anaplastic thyroid cancer (ATC) all originate from follicular cells [49, 52]. Medullary thyroid cancer (MTC) originates from C cells and often spreads to lymph nodes and distant organs [12, 38]. The different histological variants of thyroid cancer influence the patient management options and aid in determining prognosis. The follicular cell types can be categorized into differentiated thyroid carcinomas (DTC), poorly- differentiated carcinomas and undifferentiated thyroid carcinomas.

2.2.2.1 Differentiated thyroid cancers (DTC)

PTC and FTC are the main types of differentiated thyroid carcinoma. Subtypes of DTC include papillary thyroid microcarcinoma (PTMC), encapsulated papillary variant, aggressive papillary variants and subtypes of FTC such as HCC [38, 53]. PTC has a lymphatic route of spread, usually to the regional lymph nodes while FTC has haematogenous metastatic spread, commonly to distant organs such as bones and lungs [12, 51, 54]. A higher incidence of FTC is linked to iodine deficiency whereas, in iodine-sufficient areas, PTC is the most prevalent thyroid cancer. PTC tends to have a good prognosis and is less aggressive than FTC. The role of iodine in thyroid

carcinogenesis appears to be the interchange of tumour architecture from follicular to papillary carcinoma in experimental animal models [40]. Poorly-differentiated thyroid cancers such as insular carcinoma tend to be aggressive with poor prognosis.

2.2.2.2 Undifferentiated thyroid cancers

ATC falls under this class of thyroid cancers. ATC is rare, affects mostly patients above 65 years old and accounts for less than 5% of all thyroid cancers. It usually manifests as a rapidly growing palpable neck mass with distant metastases to the lungs, brain and bones [54]. ATC is very aggressive and has a highly infiltrative invasive disease process that replaces most or all of the normal thyroid parenchyma in most patients [55]. It has a poor prognosis and accounts for almost 50% of all thyroid cancer mortalities [56].

2.2.3 Genetic influence in thyroid cancers

Genetic influence has been linked to thyroid cancer occurrence. BRAF and RAS mutations and RET and TRK rearrangements have been found to be associated with PTC, while RAS mutations have also been linked to follicular PTC and FTC [51, 57]. The majority of MTC cases can be attributed to the RET oncogene mutation, particularly those associated with multiple endocrine neoplasia (MEN-II) syndromes and familial types [58]. The development and progression of poorly-differentiated and undifferentiated thyroid cancers have been attributed to RAS, BRAF and tumour suppressor gene (TP53) genetic mutations [57].

The BRAF V600E mutation is reported to frequently occur in up to 69% of PTCs; most commonly in classical PTCs and the tall-cell variant of PTC [59, 60]. Some studies reported an association between a high prevalence of the BRAF V600E mutation and high iodine levels or high concentrations of chemical elements such as iron and manganese in drinking water [61]. The BRAF V600E mutation has been linked to high mortality rates which are associated with advanced age (≥ 65 years) at diagnosis, advanced thyroid cancer stage, extrathyroidal

extension, lymph node metastases and distant metastases [62, 63]. The high mortality rates are attributed to aggressive tumorigenesis which is linked to PTC de-differentiation that causes suppression of the gene expression of the NaIS thereby resulting in radioiodine therapy resistance and the high risk of tumour recurrence [64-66]. Although the BRAF V600E mutation is highly predictive of disease recurrence with adverse clinical outcomes, it is not an exclusive poor prognostic factor as less than 15% of the PTCs that harbour the mutation have been reported to exhibit aggressive disease characteristics [61, 62, 67].

2.2.4 Risk factors, staging and prognostic factors in thyroid cancer

Certain clinical features are linked with an increased risk of thyroid cancer. These include age less than 20 or older than 60 years; a family history of DTC, MTC and MEN-II syndromes, history of Hodgkin and non-Hodgkin's lymphoma, a history of neck irradiation for head and neck malignancy and total body irradiation for bone marrow transplantation [68, 69]. Symptoms that are indicative of very high suspicion of malignancy are a rapidly growing nodule accompanied with complaints of dysphagia and dysphonia and evidence of vocal cord paralysis and cervical lymphadenopathy [70, 71]. Suspicious findings on physical examinations include a thyroid nodule that is firm on palpation, fixated onto surrounding structures, larger than 4cm, cervical lymphadenopathy and immobility of the vocal fold [72]. In some studies, cervical lymphadenopathy and vocal cord paralysis demonstrated a 100% positive predictive value for thyroid malignancy [73].

Diverse staging systems are used to classify patients into different risk groups and to guide management. These include:

- **AGES: A**ge at presentation, **G**rade of tumour, **E**xtent, **S**ize of the primary tumour
- **AMES: A**ge at presentation, **M**etastases, **E**xtent, **S**ize of the primary tumour
- **EORTC: E**uropean **O**rganisation for **R**esearch and **T**reatment of **C**ancer methodology,

- **MACIS: M**etastases, **A**ge at presentation, **C**ompleteness of surgical resection, **I**nvasion (extra-thyroidal), **S**ize, and
- **TNM: T**umour size, **N**odal metastases and distant **M**etastases [69]

The TNM system is the most commonly used system, which is applied to the three common thyroid cancers, PTC, FTC and MTC [58]. Staging aids in determining patient prognosis. The age of the patient at diagnosis, male gender, the extent of the tumour based on size, tumour histology, presence of nodal and distant metastases as well as extra-thyroidal invasion influence prognosis. Although prognosis factors are inter-dependent, differentiation and stage at diagnosis and tumour histology are strong independent risk predictors [74].

Because thyroid nodules are more common in females, the diagnosis of thyroid cancer in males is often at an advanced stage hence the male gender link to poor prognosis [75-77]. The link between the risk of thyroid cancer and advancing age remains unclear. One study indicated an increase in poor prognosis in well-DTCs at a cut-off age of 55 years [78], while another study indicated a cut-off of 45 years [38]. Age influences prognosis greatly in MTC and FTC than PTC [74]. Advanced disease as signified by a large tumour extent, locoregional invasion of organs and presence of regional and distant metastases is linked to poor prognosis. Well-DTCs usually have a favourable prognosis with PTC by having a better prognosis than FTC, while poor or undifferentiated cancers have a poor prognosis [69]. Widely-invasive FTC has a poor prognosis due to gross extrathyroidal infiltration and significant vascular invasion [51]. DTCs tend to be less aggressive with a good 10-year survival rate and only about 10% of patients succumb to cancer; however, there is a 5 to 20% risk of recurrence as well as up to 15% risk of distal metastases [58, 79]. MTC is minimally invasive and has a 5-year survival rate of up to 85%, however, the prognosis is poor with advanced stage, distant metastases or metastatic disease [74]. ATC has a poor prognosis due to its extremely aggressive nature and diagnosis at advanced

stages [79]. Ultimately, the type of thyroid cancer and stage at diagnosis influence the management approach and prognosis outcomes.

2.3 Laboratory tests for thyroid cancer

Thyroid cancer diagnosis involves laboratory tests and diagnostic imaging. In laboratory testing, blood analyses are used for thyroid tests to screen thyroid function and for the assessment of appropriate treatment for DTC [42]. In patients with palpable or incidental thyroid nodules, diagnostic testing initially involves the assessment of risk factors and the measurement of serum thyroid-stimulating hormone (TSH) to ascertain hypo- or hyperthyroidism [80]. A radionuclide scan or thyroid scintigraphy is conducted in the event of a low serum TSH finding.

2.3.1 Thyroid function tests

Thyroid function tests include measurement of serum TSH, serum T_3 (triiodothyronine) and T_4 (thyroxine), thyroid antibodies and thyroglobulin levels. The need for the performance of the tests is dependent on other factors such as symptoms, risk factors, patient age, gender, and the presence of comorbidities. The use of thyroid function tests should be patient-specific except for serum TSH which ought to be measured in all cases of thyroid nodular disease [42, 81].

2.3.1.1 Serum TSH and serum fT4

This is the recommended function test in all patients with thyroid nodules. The accepted normal range for serum TSH is about 0.4 to 4.0 mIU/L despite there being varying opinions on the appropriate upper normal limit [42]. An elevated TSH within the normal upper limit or greater than 4.5 mIU/L is linked with increased thyroid cancer risk and an advanced stage of malignancy [81, 82]. However, the reason for the increased risk of thyroid cancer with upper baseline TSH levels remains unexplained. Some studies alluded to the role of TSH as a growth factor of thyroid follicular cells, suggesting that TSH may promote the growth and

aggressiveness of already present thyroid neoplasms and not necessarily initiate carcinogenesis [83, 84]. Furthermore, since DTCs express TSH receptors, TSH may not have an independent oncogenic role but may have a catalystic role in thyroid tumorigenesis when other growth factors and oncogenes are present [66, 85].

When serum TSH is low, thyroid scintigraphy is indicated for the assessment of thyroid nodule functionality as most hyper-functioning nodules are rarely malignant and require no further cytology assessment [68]. However, thyroid function assessment by measurement of serum TSH is most accurate when there is no co-existent pituitary gland or hypothalamic pathology and no drugs are affecting the thyroid. Serum fT_4 is the alternate test to estimate thyroid hormone levels in patients on anti-thyroid drugs or for the assessment of the effectiveness of radio-iodine therapy and surgery when TSH levels are unstable [42].

2.3.1.2 Serum thyroglobulin (Tg) and thyroid antibodies

Serum thyroglobulin (Tg) measurement is not recommended in the routine for primary evaluation of thyroid nodules as levels tend to be elevated in many thyroid pathologies. However, the test has relevance in the follow-up, monitoring of PTC, and detecting post-thyroidectomy recurrence [72, 81]. Optimal post-thyroidectomy Tg levels should be < 1ng/ml, with values <0.2ng/ml indicating a good response to therapy and the low likelihood of recurrence [71, 86]. The prevalence of thyroid antibodies is high even in normal thyroid glands; hence they are also not routinely measured in primary thyroid nodule assessment [80, 87].

2.3.2 Serum calcitonin and carcino-embryonic antigen (CEA)

Serum calcitonin and CEA measurements are indicated in patients with a familial history of MTC, MEN-II syndromes or hyperparathyroidism because elevated serum calcitonin levels (above 100pg/ml) correspond with a high risk of medullary thyroid cancer [72, 81]. When the test is performed for the staging, surveillance and prognosis assessment of MTC, calcitonin levels

above 1000pg/ml are consistent with distant metastases [88]. Divergent perspectives suggest that serum calcitonin levels are indefinite in thyroid cancer diagnosis due to the high false-positive rates in hypercalcemia and neuroendocrine tumours and high false-negative rates in rare non-calcitonin MTCs [72, 80]. Serum calcium and parathyroid hormone (PTH) tests are further indicated in patients at risk of MTCs to exclude hyperparathyroidism [89].

2.3.3 Molecular markers

This emergent technique relies on the recognition of gene mutations that correspond to certain types of thyroid cancers. Genetic markers such as RET/PTC, RAS, BRAF and the protein marker Galectin-3 are used as molecular markers. Molecular markers are proposed to be more helpful in the diagnosis of thyroid nodules with indeterminate cytology [80]. The presence of RET/PTC rearrangements or BRAF mutations in samples of indeterminate cytology has been suggested to be highly specific for thyroid cancer, whereas RAS mutations have a malignancy risk of up to 87% [69, 72]. One study demonstrated an increased specificity of fine needle aspiration cytology (FNAC) from 36% to 95% when FNAC and BRAF mutation assessments were combined [90]. However, since the technique is still relatively new, current molecular marker tests cannot give a conclusive thyroid cancer diagnosis in nodules with indeterminate cytology. Current molecular marker tests can miss up to 33% of cancers and have low sensitivity and negative predictive values [61, 91]. Furthermore, the limitations of the tests are that they are very expensive, not widely available and their potential utility is limited to classic papillary cancer and not follicular cancers [80, 92]. The current utility of molecular markers can therefore be suggested supplementary to FNAC and diagnostic imaging approaches depending on patient history, availability of facilities and experienced personnel conducting the tests.

2.3.4 Fine needle aspiration cytology (FNAC)

Fine needle aspiration cytology (FNAC) is the pre-surgical gold standard test for the evaluation of thyroid nodules. This minimally invasive technique can be done under ultrasound guidance or using palpation to obtain aspirates of epithelial cells from a thyroid lesion. Ultrasound-guided FNAC is the most recommended approach as it has lower non-diagnostic rates of up to 7% while non-diagnostic rates in the palpation-guided approach can be more than 30% [93]. Furthermore, ultrasound-guided FNAC has low false positive rates and is valuable for posterior, deep and impalpable nodules [81]. Clinicians conducting the biopsies should have sufficient knowledge of ultrasound features of thyroid lesions for the adequate FNAC selection of thyroid nodules.

The factors that influence the indication for FNAC include the patient's history, the sonographic findings and the size of the nodule [77]. FNAC is generally recommended for all nodules meeting any of the following criteria:

- greater than 1cm in diameter with highly suspicious sonographic features (solid, hypoechoic, microcalcifications, taller-than wide),
- less than 1cm with clinical risk factors highly suggestive of thyroid malignancy (family history of thyroid cancer, history of neck irradiation) and abnormal cervical lymph nodes and a suggestive sonographic feature, and
- any size nodule with cervical lymph node involvement and/or extrathyroidal extension [70, 81].

For thyroid nodules larger than 3cm, there can be an increased sampling error rate in FNAC of over 30% hence diagnostic lobectomy may be considered given other risk factors and symptoms [72].

There are several systems of classifying the results of FNAC. However, the general system constitutes four categories; benign, malignant, indeterminate and non-diagnostic [44]. Among the currently available systems, The Bethesda System for Reporting Thyroid Cytology (BSRTC) is the most common globally as it is considered standardized, uniform, and consistent [94, 95]. The BSRTC has six categories:

1. non-diagnostic,
2. benign,
3. follicular lesion or atypia of undetermined significance (FLUS or AUS),
4. follicular neoplasm (FN) or suspicious for a follicular neoplasm (SFN) (including Hurthle cell lesions),
5. suspicious for malignancy, and
6. malignancy [96, 97].

The risk of malignancy varies incrementally across the different categories from the lowest to the highest level and this determines the choice of management. While categories 2 and 6 are often determinate diagnoses, categories 1, 3, 4 and 5 tend to be inconclusive. Some clinicians exclude category 5 from the indeterminate category due to its high risk of malignancy of up to above 75% [98, 99]. About 10 to 15% of FNAC results are non-diagnostic (category 1), often due to insufficient cell samples for definitive diagnosis, whereas 10 to 30% are indeterminate due to equivocal cytology findings , i.e. categories 3, 4 and 5 [100]. A category 1 diagnosis will require repeat biopsy and sampling. However, the cytologically equivocal findings present a major diagnosis challenge particularly FLUS/AUS and FN/SFN that have indistinct features which overlap between benign and malignant follicular lesions [68, 70]. Furthermore, borderline tumours such as the non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) are classified under the indeterminate category as they cannot be reliably distinguished from PTC and FTC cytologically [101].

Most cytologically equivocal nodules often undergo repeat FNAC and subsequent diagnostic surgery for definitive diagnosis yet some histopathology analyses have shown that 70 to 80% of these tend to benign [6, 102]. This implies that some patients with benign disease are unnecessarily subjected to risks of surgery such as laryngeal nerve paralysis and post-surgery hypothyroidism and hypoparathyroidism which result in a lifetime of thyroid hormone replacement therapy [91, 101]. One study demonstrated 18% surgery-related complications predominated by hypoparathyroidism (12%) after total thyroidectomy of follicular neoplasms [103]. A recent meta-analysis showed that resection rates of cytologically equivocal nodules were higher in Western practices (51.3%), while Asian practices were slightly lower (37.6%) [104]. It is alluded that most Asian practices treat cytologically equivocal nodules more conservatively such as the management of NIFTP as benign follicular adenoma whereas in some Western practices this can be regarded as a low-risk tumour and treated as a malignant nodule [105]. Some studies have suggested that some cytologically equivocal nodules may not require total thyroidectomy or subsequent radio-iodine treatment but follow-up and/or conservative lobectomy surgery may suffice thereby limiting post-surgery complications [103, 105, 106]. However, the challenge that persists is that there is no conclusive diagnosis method for these nodules before the diagnostic surgery, and in some cases, the best management approach can only be determined based on the surgical findings.

The diagnostic performance of FNAC in thyroid nodule diagnosis varies in different studies. Some studies showed that sensitivity ranges from 83 to 98% [107], specificity ranges from 70 to 92% [107-109] and diagnostic accuracy can range between 84 to 98% [109-111]. The range of false-negative rates was from 1 to 19% while false-positive rates were from 6 to 8% in one study [109]. Another study suggested that the estimated average rates of sensitivity and specificity were 83% and 92% respectively, while both the false positive and false negative rates were about 5% [92]. Similarly, a recent meta-analysis demonstrated a slightly lower pooled sensitivity (72%)

in comparison with the specificity (99%) with a positive likelihood ratio (PLR) of 41.71 [112]. The differences of these diagnostic performance values from these different studies can be attributed to the different study designs in addition to the reliance of FNAC on the accuracy of sampling techniques, sufficient specimen and the skill of the cytopathologists.

FNAC is limited in distinguishing follicular adenoma from follicular carcinoma and diagnosing follicular PTC. Another major limitation is the occasional high rate of inadequate or non-diagnostic samples [69]. Non-diagnostic and inadequate samples may be a consequence of the limited experience of the cytopathologist, the poor nature and low quantity of the aspiration and poor specimen preparation [70, 113]. Diagnosis results from FNAC contribute to the final management decision in thyroid nodule diagnosis hence there is a need for accurate and reliable results.

2.4 Major diagnostic imaging modalities in thyroid cancer

2.4.1 Ultrasound imaging in thyroid cancer

Ultrasonography is the primary imaging modality in the assessment of thyroid nodules. It is mainly used in the initial evaluation of thyroid nodular disease due to its high spatial resolution, excellent temporal resolution, non-ionising nature, and ease of use. In clinical practice, a thyroid ultrasound examination is done routinely for all patients with suspected or known thyroid pathology. Ultrasonography imaging helps to confirm the presence of a nodule, demonstrate features with/without suspicion of malignancy and cervical lymph nodes [80]. Furthermore, it aids in the decision-making for triaging thyroid nodules for either FNAC or follow-up, guides FNAC and is useful in the pre-operative staging and post-operative evaluation of thyroid cancer [69, 114].

2.4.1.1 Grey scale assessment

There are characteristic sonographic features observed using different ultrasound modes that are suggestive of a benign or malignant disease process. A normal thyroid gland has homogenous hyperechoic echotexture relative to the adjacent muscles (i.e., strap muscles – sternohyoid and sternothyroid; sternocleidomastoid), and clear margin outlines (*Figure 2:2* and *Figure* 2:3). Any deviation from the normal appearance is usually indicative of a thyroid nodular disease process.

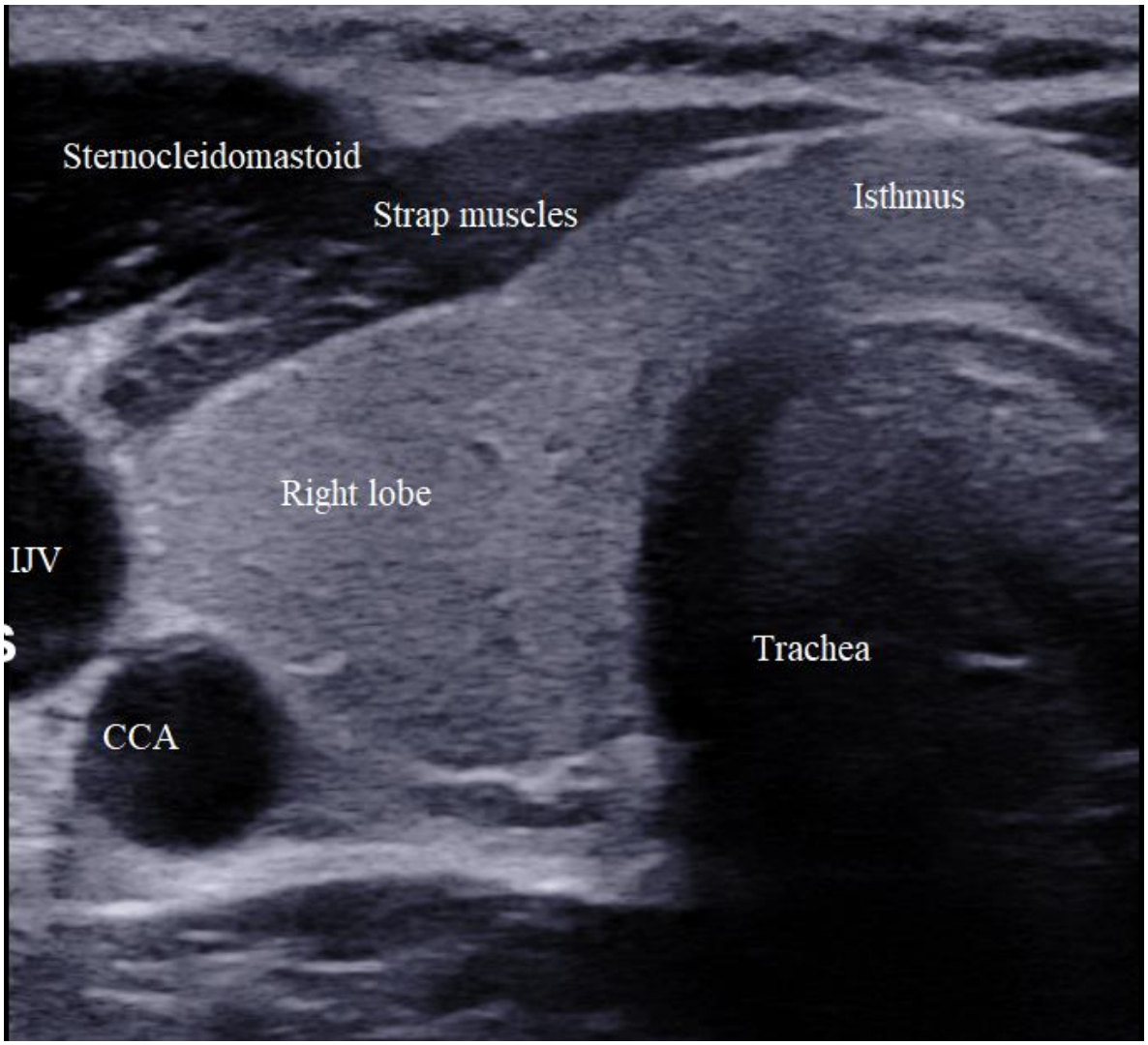


Figure 2:2 A transverse grey scale ultrasound scan of a normal right lobe of the thyroid gland. The adjacent structures such as the trachea, strap muscles, sternocleidomastoid muscle, the right internal jugular vein (IJV) and the right common carotid artery (CCA) are shown.

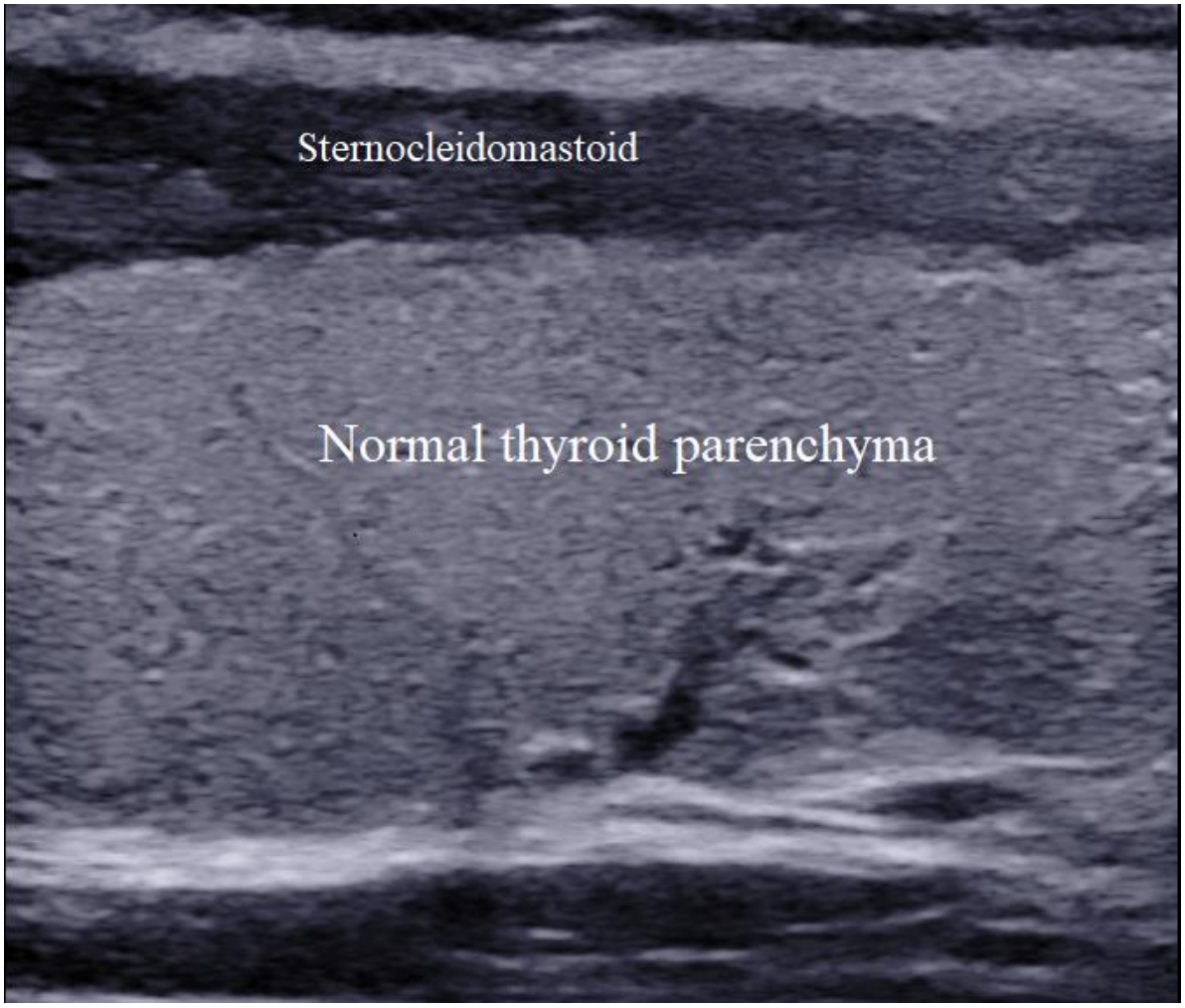


Figure 2:3 A longitudinal grey scale ultrasound scan of a normal right lobe of the thyroid gland.

2.4.1.1.1 Ultrasound features suggestive of benign and malignant thyroid disease

The common ultrasound features that have been established to be indicative of a benign thyroid nodule include a completely cystic nodule, spongiform appearance, peripheral halo sign (*Figure 2:4),* regular well-defined margins, peripheral vascularity, comet tail sign and coarse calcifications [9, 10, 115]. Sonographic features that are highly suggestive of malignancy include a purely solid or mostly solid nodule, hypoechogenicity relative to the strap muscles, a taller than wide nodule (*Figure 2:5*), absent halo sign, irregular margins and/or shape, microcalcifications and the presence of abnormal lymph nodes [4, 19, 81, 116-118]. When there are no suspicious sonographic findings in the thyroid, the sole finding of abnormal lymph nodes is highly indicative of malignancy [69, 70, 114]. Cervical lymphadenopathy can be detected using pre-operative ultrasound in up to 30% of patients, thereby altering the surgical management and improving the RAI therapy approach [119].

Of the features that most authors agreed are highly predictive of malignancy, the most common is marked hypoechogenicity in a solid nodule which has an average sensitivity of about 94% [12, 25, 120]. In one study marked hypoechogenicity was most predictive of malignancy than other suspicious ultrasound features with the highest odds ratio > 15 [121]. Similarly, some studies showed that more specifically, a markedly-hypoechoic nodule is highly suggestive of malignancy than just a hypoechoic nodule [121, 122]. However, in another study, moderate hypoechogenicity had a higher malignancy risk (71.2%) than marked hypoechogenicity (60.5%) in nodules with suspicious features [123]. Furthermore, another study showed a lower sensitivity of about 41% and the presence of this feature in up to 55% of benign nodules [19]. Similarly, Smith-Bindman *et al*., [124] demonstrated that microcalcifications had a higher odds ratio (11.6) than hypoechogenicity (2.9); however, PTCs constituted the most thyroid cancers in that study (81%). However, some studies found that multiple suspicious sonographic features in a sole nodule raised the risk and positive prediction of malignancy [4, 125]. In one

study, the finding of a solid hypoechoic nodule with irregular margins, intranodular vascularity or microcalcifications had a positive predictive value of 87% in identifying thyroid cancers [126].

–

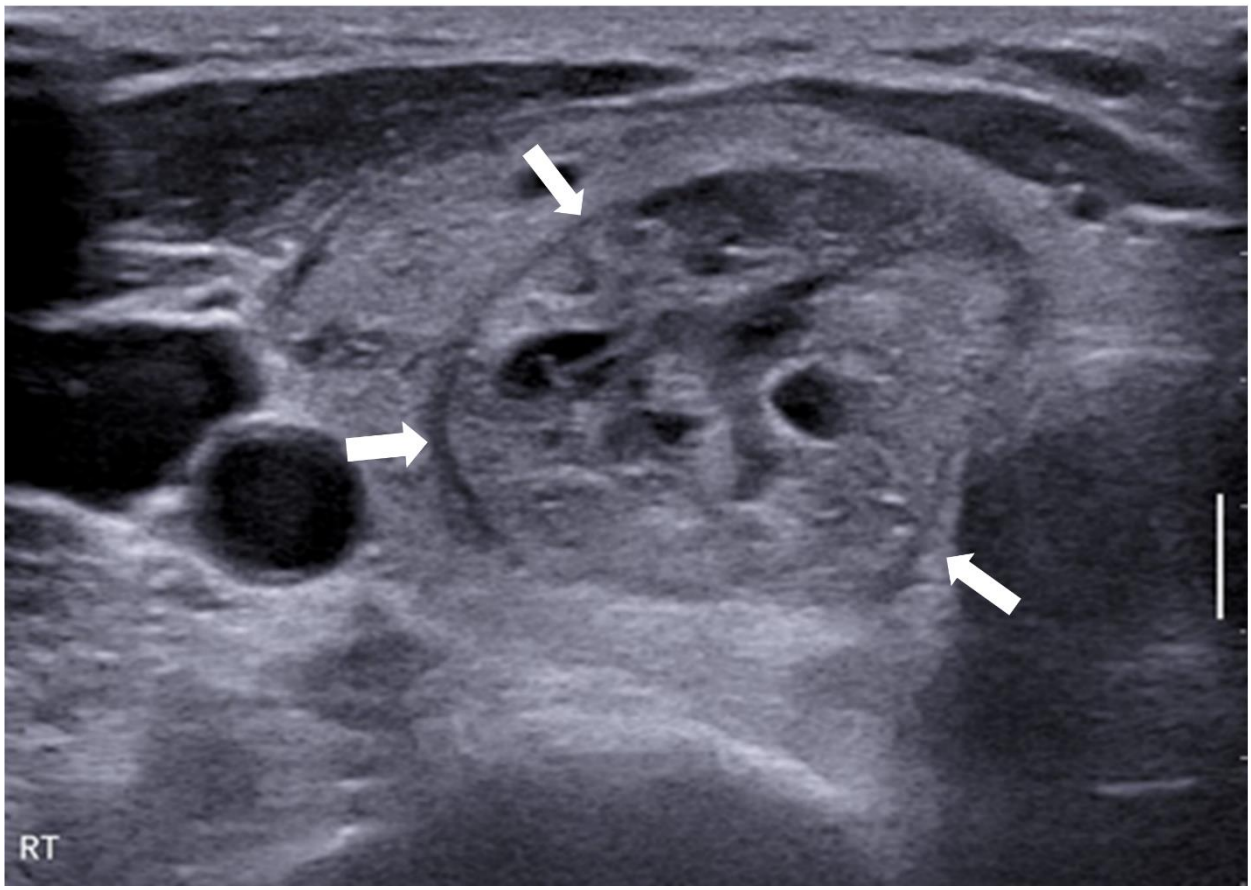


Figure 2:4 A typical benign thyroid nodule illustrating a spongiform appearance and complete peripheral halo sign (arrows).

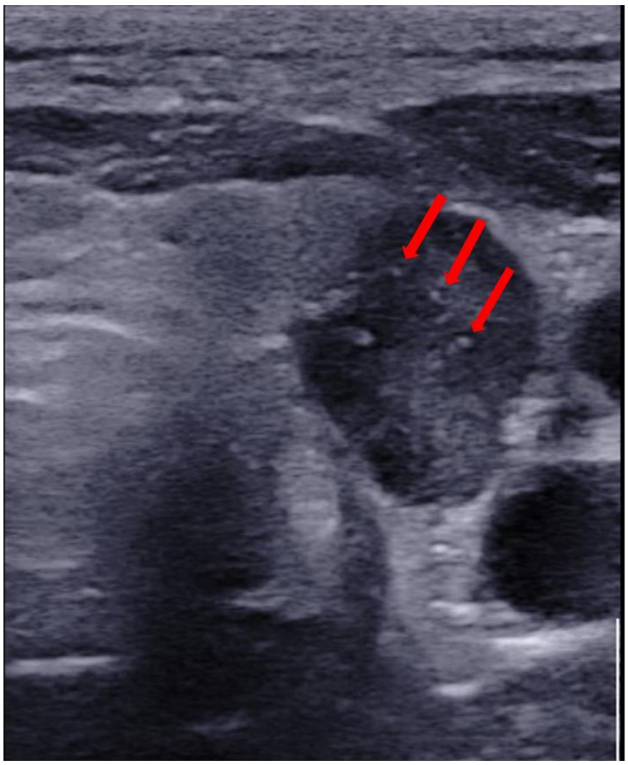


Figure 2:5 A left thyroid lobe transverse sonogram demonstrating a malignant thyroid nodule. This was confirmed as papillary thyroid carcinoma (PTC) in cytology and histopathology in a 73-year-old female patient. The nodule was solid, taller than wide, markedly hypoechoic and had microcalcifications (arrows).

Grey scale ultrasound often results in indefinite results in thyroid nodules with equivocal cytology due to the overlap in features between benign and malignant neoplasms in this category. Nonetheless, some studies suggested that microcalcifications and fast nodular growth are highly predictive of malignancy while others found hypoechogenicity and the taller than wide ratio most predictive [127, 128]. This variability in results is indicative of the present persistent challenge in the accurate diagnosis of thyroid nodules with equivocal cytology as these nodules sometimes have atypical features.

Ultrasound has the limitation of being highly operator-dependent which can yield different results among different sonographers and clinicians. Additionally, there is an overlap of some features that can be found in both malignant and benign nodules and no single sonographic feature is highly diagnostic on its own [20, 116, 129]. Furthermore, ultrasound cannot accurately differentiate between the different types of cancer as well as differentiating follicular adenoma and follicular carcinoma[71, 81, 130]. The diagnosis of follicular neoplasms can only be made on histopathology based on vascular or capsular invasion [131]. The typical features that are suggestive of malignancy such as microcalcifications are seen mostly in PTCs than in other thyroid cancers. Therefore, there are still some nodules in which ultrasound results may be inconclusive and these may have to undergo FNAC and surgery for definitive diagnosis.

2.4.1.1.2 Ultrasound malignancy-risk stratification systems

Risk-stratification systems (RSS), also known as Thyroid Imaging Reporting and Data System (TIRADS) guidelines, emerged as an effort to standardise thyroid nodule malignancy-risk stratification and improve diagnostic accuracy. These guidelines assist with the differentiation of thyroid nodules by malignancy-risk stratification and FNAC recommendation based on several predictive sonographic features. While the different TIRADS use the same standard ultrasound features for malignancy risk assessment, i.e. echogenicity, composition,

calcifications, margins and taller than width ratio and shape, they may differ in the sub-classifications and risk assignment criterion of the features, while a few incorporate the assessment of vascularity and elasticity [132]. As a result, different RSS can be classified as score-based or pattern-based through score assignment and calculating the sum of scores of all present ultrasound features or applying high-risk assignment for multiple features predictive of malignancy, respectively [133]. The score-based criteria assign higher scores (usually on a scale of 0-3) to ultrasound features that have been established to have a high prediction of malignancy independently.

The most commonly used TIRADS guidelines include those from the American Thyroid Association (ATA), American College of Radiology (ACR), the American Association of Clinical Endocrinologists, the American College of Endocrinology, and the Associazione Medici Endocrinologi (AACE/ACE/AME referred to as AACE from here-on), European Thyroid Association (EU), British Thyroid Association (BTA) and the Korean Society of Thyroid Radiology/Korean Thyroid Association (KSThR/KTA) [70, 71, 89, 134-136]. Most TIRADS consist of 5-6 categories in incremental risk order from benign to most suspicious of malignancy, except for AACE which has only 3 categories (low risk, intermediate-risk, and high risk). Studies on the different guidelines have shown variable results; however, most current studies assert that ACR is the best guideline for reducing the unnecessary biopsy rate [137-140]. Schenke et al., [141] demonstrated the highest negative predictive value of 100% with ACR but the highest positive predictive value and accuracy (70.5% and 77.2%, respectively) were achieved using Kwak TIRADS [142]. Some of the variabilities in the TIRADS that yielded the highest sensitivity from different studies are Kwak (98.3%), ATA (95%), EU (94.7%) and KSThR (91.7%) [143-146]. The study design, the choice of TIRADS for comparison and the criteria for final diagnosis (histopathology vs FNAC) likely contribute to the divergences among other different factors. Furthermore, although most TIRADS may be agreeable on the common

features, there are some nodules that other TIRADS may fail to classify such as nodules with mixed cystic composition with isoechogenicity [70, 71, 89]. These areas of divergences present diagnosis challenges and the need for more studies evaluating the sole and combined use of the different TIRADS in ultrasound thyroid nodule diagnosis.

The utility of TIRADS in the sonographic diagnosis of thyroid nodules with equivocal cytology remains unclear. One study demonstrated high negative predictive values with ATA and ACR of up to 100% while other studies using different TIRADS found minimal to no value in ruling out malignancy [147-150]. The differences in malignancy-risk criteria and estimation even among shared common predictive features explains the conflicting results in different studies and therefore no consensus on the best TIRADS to use to date. Furthermore, TIRADS guidelines generally have a major drawback of low specificity [121, 136, 151]. Due to the current limitations, the challenge of achieving a definite thyroid nodule diagnosis solely based on TIRADS persists.

2.4.1.2 Vascularity assessment

The high metabolic activity in cancerous cells results in an increase in thyroid nodule vascularity due to increased blood supply. This is attributed to angiogenesis - cancer cells invading and growing in regions deficient in blood vessels; thereby creating new blood vessels (neovascularization) for survival, growth and proliferation [152]. In ultrasound, blood flow detection in a vascularized area is based on the Doppler effect principle - the change in the frequency of sound waves secondary to the motion of blood relative to the direction of incident ultrasound beam [153]. The resultant change in the frequency between the emitted and reflected ultrasound waves is the Doppler shift; whereby a positive Doppler shift with the red blood cells moving towards the transducer is commonly denoted by a red colour in a colour scale, a negative Doppler shift with the cells moving away from the transducer is usually denoted by a

blue colour. The directional and blood velocity-dependent colour flow is known as colour Doppler whereas its amplitude dependent counterpart is known as power Doppler.

2.4.1.2.1 Colour and Power Doppler in thyroid nodule vascularity assessment.

Colour Doppler ultrasound (CDUS) and Power Doppler ultrasound (PDUS) are instrumental in the assessment of the vascularity of thyroid nodules. CDUS can be used to demonstrate direction and presence of blood flow, as well as flow velocity measurements, whilst PDUS can demonstrate the maximum presence of flow without flow velocity measurements. PDUS is more sensitive to slow blood flow with better detail of the total amount of flow than CDUS, as it depends on the number of Doppler shifts and is not prone to aliasing due to the lack of directional flow information when compared to CDUS [154]. Therefore, PDUS is suggested to be more accurate in the characterization of thyroid nodules. There are diverse opinions regarding the role of vascularity assessment in predicting thyroid nodule malignancy. Predominant central vascularity is suggested to be indicative of malignancy whereas peripheral vascularity is associated with the benign disease process in characterising thyroid nodules (*Figure 2:6* and *Figure 2:7*) [12, 20, 116, 155]. Studies using CDUS in vascularity assessment have demonstrated that central vascularity is associated with malignancy with high sensitivity (>70%), high specificity (>90%) and odds ratio (>10) [126, 156, 157] Furthermore, a recent meta-analysis showed that CDUS had a pooled sensitivity of 74%, pooled specificity of 70% and an accuracy of 0.78 based on the area under the curve analysis (AUC) in the characterization of thyroid nodules [158]. Contrarily, some studies have indicated the vascularity assessment with CDUS in diagnosing malignancy had no additional value with some studies showing low sensitivity (30%) even with high specificity (>80%) [130, 159, 160]. PDUS is suggested to be highly sensitive in detecting slow blood flow and one study demonstrated a high odds ratio of 219 in predicting malignancy using a predominantly central vascularization pattern [13, 161]. In another study, the density of central vascularity with 3D PDUS was the main independent risk factor for thyroid malignancy

with a sensitivity of 75 to 81% and a specificity of 49 to 56% [162]. Although 3D ultrasound has the advantage of volumetric analysis of vascularity of vessels in a 360-degree rotation, its limited use in neck ultrasound is attributed to high computational needs and reconstruction artefacts [163]. Contrarily, some studies indicated that PDUS vascularity assessment was not useful even in combination with grey scale ultrasound features since central vascularity may also be observed in benign nodules while some malignant nodules may have absent vascularity [159, 164-166].

The conflicting findings regarding the role of thyroid nodule vascularity assessment in predicting malignancy may be attributed to the differences in the qualitative subjective methods used in assessing thyroid nodule vascularity. The common qualitative methods of vascularity assessment can be based on the classification of the pattern or extent of blood flow into the following:

Three–type classification – Type 0 = no vascularity; Type 1 = peripheral vascularity and Type 2 = central vascularity with or without peripheral vascularity [126, 167],

Four–type classification – Type 1 = absent vascularity; Type 2 = peripheral vascularity only; Type 3 = mild intranodular vascularity (<50%) with or without peripheral vascularity and Type 4 = marked intranodular vascularity (>50%) with or without peripheral vascularity [136, 157], and

Five-type classification – Type I =absence of blood flow; Type II = exclusively peripheral blood flow; Type III = predominantly peripheral flow, Type IV = predominantly marked central blood flow and Type V = exclusively central blood flow [161].

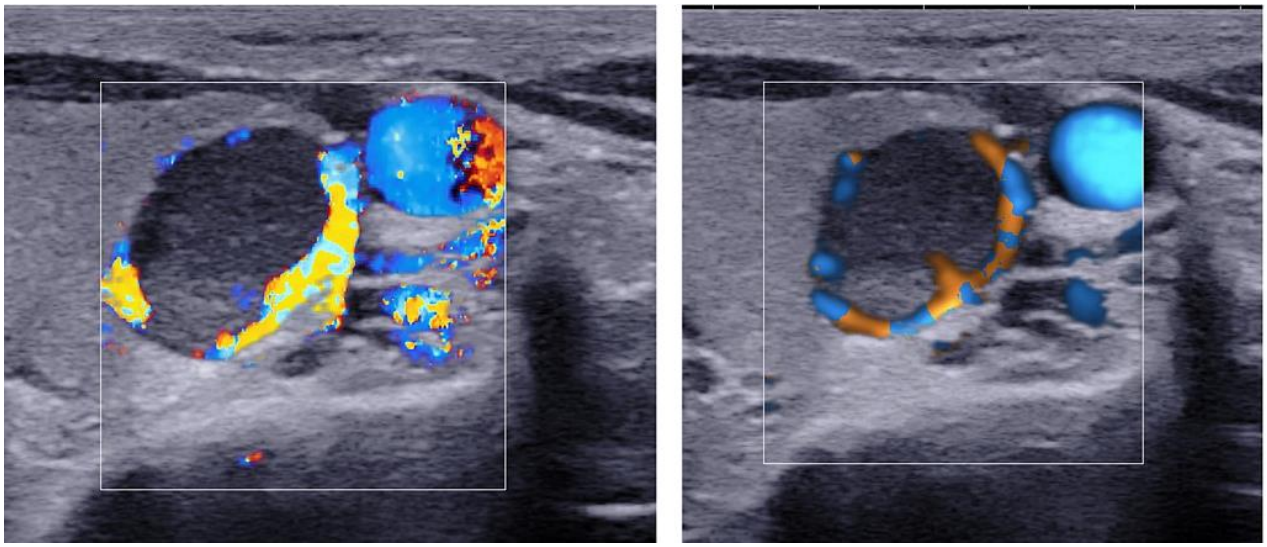


Figure 2:6 Transverse scan sonograms illustrating peripheral vascularity in CDUS (left) and PDUS (right) in a benign nodule on the left thyroid lobe of a 40-year-old female patient. The nodule was confirmed to be a follicular adenoma in histopathology.

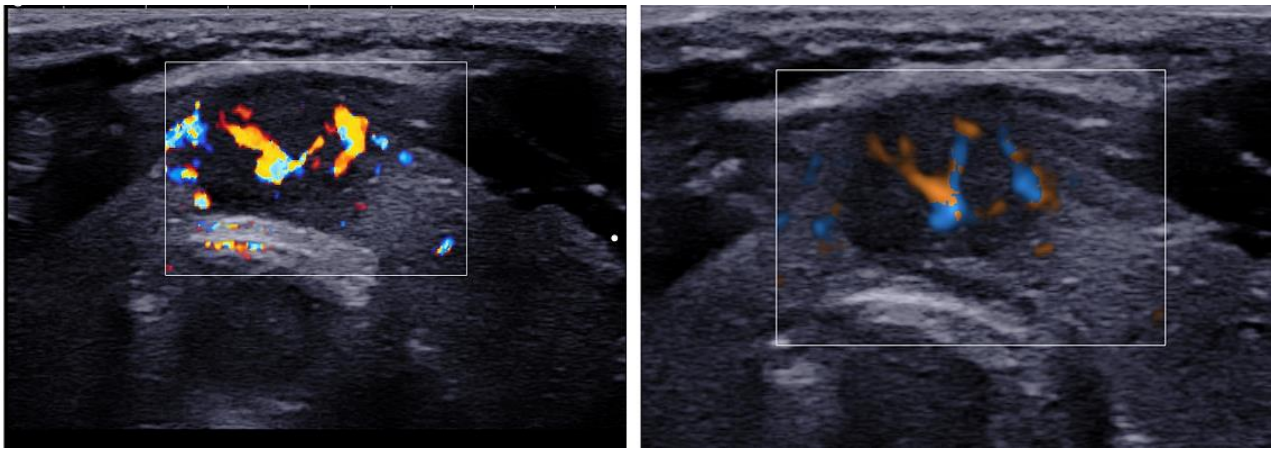


Figure 2:7. Transverse scan CDUS (left) and PDUS (right) sonograms illustrating central vascularity in a malignant nodule. The nodule was located on the isthmus of thyroid gland of a 60-year-old female patient with histopathologically confirmed papillary thyroid carcinoma.

Currently, there are few studies on the quantitative assessment of thyroid nodule vascularity. Recent studies based on the quantitative regional analysis of thyroid nodule vascularity and cervical lymph nodes have shown that quantitative determination of vascularity when used with other grey scale ultrasound features, has increased reliability in predicting malignancy in contrast with subjective means of determining thyroid nodule vascularity [23, 168-170]. In a previous study, it was demonstrated that the addition of regional vascularity indices (VI) to grey scale ultrasound assessment improved the diagnostic accuracy from 58.6% to 79.3% and specificity from 46.4% to 83.3% [23]. Contrarily, Yoon *et al*.,[171] indicated that quantitative VIs based on PDUS offered no additional value to grey scale assessment. Although quantitative thyroid nodule vascularity methods continue to evolve, CDUS and PDUS are still deficient in assessing microvascularity patterns that are most likely to arise due to angiogenesis in cancer.

2.4.1.2.2 Micro-vascularity imaging in thyroid cancer

Novel thyroid ultrasound microvascular imaging techniques have recently emerged. One such technique is superb microvascular imaging (SMI) which uses an algorithm that suppresses clutter and tissue motion to show microvascular blood flow [172]. SMI has been demonstrated to be superior to CDUS and PDUS in depicting microvascular flow, vessel branching and low blood velocity [172]. SMI has been established to be more sensitive than CDUS in the detection of microvascular blood flow in breast lesions, with a diagnostic performance comparable to that of contrast enhanced ultrasound (CEUS) imaging [173, 174]. Concurrently, studies on thyroid nodules have also demonstrated that SMI has higher accuracy than CDUS and PDUS and comparable accuracy to CEUS in diagnosing malignancy [175, 176]. A few studies involving SMI have been done; however, these studies have yielded conflicting results. In one study, thyroid intranodular vascularity on SMI had high sensitivity and specificity (75.9% and 91.2% respectively) while in combination with grey scale ultrasound features the diagnostic performance was better than any sole sonographic feature [177]. Similarly, in another study multi-modal imaging involving grey scale ultrasound, SMI and elastography improved the diagnostic accuracy of thyroid nodules with a moderate suspicion of malignancy with a sensitivity of 94.1% and specificity of 87.2% [178]. Contrarily, a similar study showed no improvement in the diagnostic performance of grey scale ultrasound in combination with elastography and Doppler ultrasound modes including SMI [179]. The difference in results may be attributed to differences in study methodology and sample sizes; however, all these studies used qualitative methods to analyse vascularity. Qualitative analysis in SMI thyroid nodule vascularity assessment is based on similar classifications as PDUS and CDUS or the number of vessel branches whereby 2 blood vessels are assigned as score 1 while >2 are score 2 [176, 177].

Another recent innovation in microvascular imaging is Angio Planewave Ultrasensitive (AngioPLUS) imaging. AngioPLUS uses spatiotemporal filtering to increase the

differentiation of low blood flow from tissue motion which makes it a highly sensitive ultrafast Doppler imaging technique [21]. Similar to SMI, it is purported to have superior low blow flow detection when compared with CDUS and PDUS modes (*Figure 2:8).* Owing to the high resolution and sensitivity in microvascular flow detection, AngioPLUS's diagnostic performance can be speculated to be comparable to that of CEUS. AngioPLUS has been suggested to be effective in differentiating parathyroid lesions from other lesions, however, there is paucity in the literature on the utility of AngioPLUS in thyroid imaging [22]. Nonetheless, unpublished data shows that the combination of AngioPLUS with CDUS and PDUS improves the detection of micro-vascularity in thyroid parenchyma [10]. There is currently a lack of studies that have evaluated the diagnostic performance of AngioPLUS in the differentiation of benign and malignant thyroid nodules. Furthermore, the value of microvascularity imaging for follicular neoplasms and thyroid nodules with equivocal FNAC results that have increased vascularity has not been extensively explored. Some studies have suggested the assessment of vascularity in these nodules may be potentially helpful in limiting unnecessary thyroidectomy surgeries [180]. Comparative qualitative and quantitative evaluation of thyroid microvascularity may help to determine the value of vascularity assessment in thyroid cancer imaging since current studies have shown discordant findings.

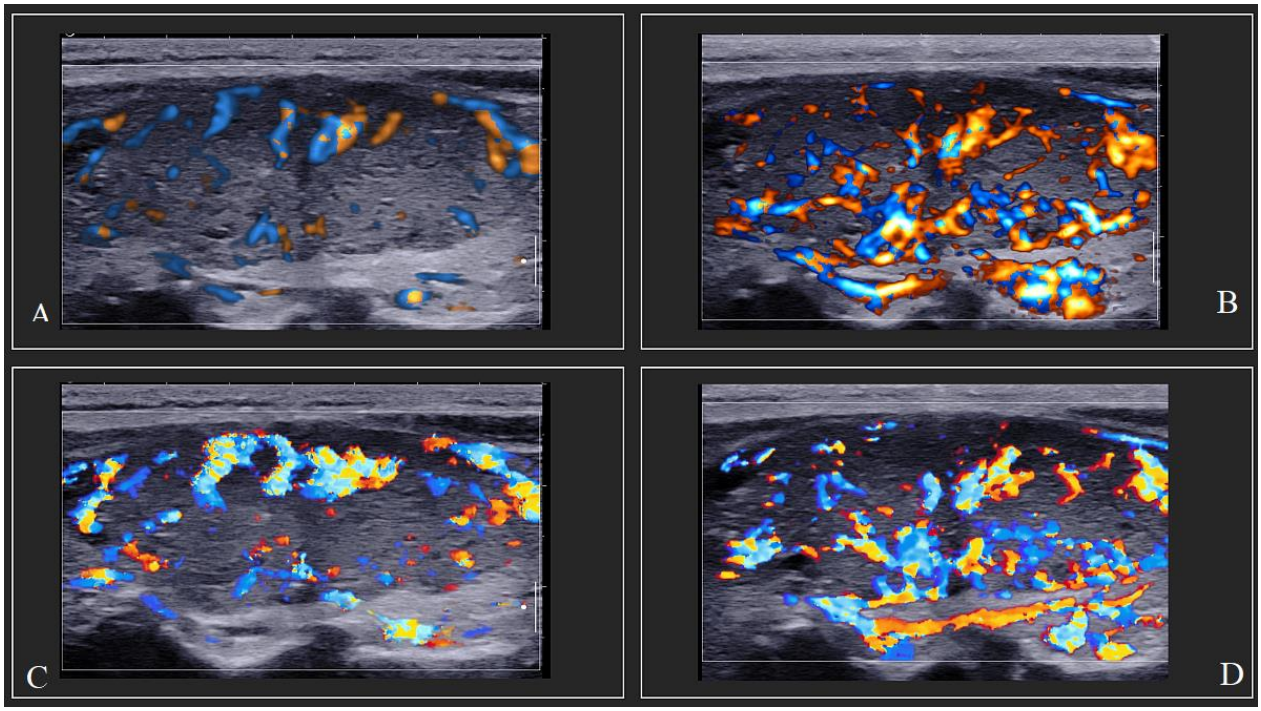


Figure 2:8 Longitudinal views of the illustration of vascularity detection using: A- PDUS, B- PDUS with AngioPLUS, C- CDUS and D- CDUS with AngioPLUS.

The AngioPLUS modes demonstrated dense microvascular flow in this case of papillary thyroid carcinoma in a 45-year-old female patient.

2.4.1.3 Elasticity assessment

Tissue elasticity is one of the physical parameters that are mostly affected by the pathophysiological process [24]. Pathophysiological processes in carcinogenesis can cause tissues to lose their elasticity resulting in malignant lesions being stiffer than benign lesions [181]. In PTCs, the predominant thyroid cancers, the increased tissue stiffness is attributed to psammoma bodies and the increased cellular compaction arising from the infiltrative nature of cancer [25]. Sonoelastography is a non-invasive ultrasound imaging technique that is used to assess soft tissue stiffness including thyroid nodule. It is based on determining tissue elasticity by inducing tissue displacement/deformation using extrinsic or intrinsic pressure and then quantifying the result as the ratio of applied pressure to strain or Young's modulus [181]. The two major methods of elastography currently used in the clinical diagnosis of thyroid nodules are strain elastography (SE) and shear wave elastography (SWE). SE involves the application of mechanical stress through the extrinsic application of transducer pressure or intrinsic pressure application using carotid pulsations and depicts results as strain ratios and colour maps to demonstrate the degree of stiffness (*Figure 2:9*) [26, 182]. SWE uses the intrinsic application of acoustic radiation impulses to generate shear waves that traverse perpendicular to the tissue resulting in the quantification of tissue stiffness as a measure of the velocity of the shear waves in quantitative elasticity colour maps measured in shear wave speed (m/s) or kilopascals (kPa) [29, 183].

In thyroid ultrasound, SE is regarded as a qualitative or semi-quantitative approach that uses strain ratios and colour maps, usually from a scale of 1 to 5, (1 representing uniformly soft tissue while 5 represents uniformly hard tissue); and red to blue (or vice versa) incrementally representing tissue stiffness for colour maps respectively [26]. In SE, the strain ratio is a relative measure of the strain of the thyroid nodule to that of the surrounding normal thyroid parenchyma whereby a ratio of > 1 is indicative of increased stiffness [181]. In SWE, the Young's

modulus is an absolute measure of stiffness whereby a high value (in kPa or m/s) is indicative of increased tissue stiffness which is suggestive of malignancy (*Figure 2:10*). Because the stiffness value obtained in SWE is an absolute measurement and quantitative, it is therefore regarded as more accurate, reliable and free from observer bias.

Some studies have indicated that SWE is more reliable in predicting malignancy in comparison to SE, while other studies found it either less superior or resulting in no difference [184-186]. The diagnostic accuracy of SWE in one study was excellent with an AUC of 0.92 [184]. However, another study demonstrated a lower diagnostic performance than SE with a sensitivity of 78.7% compared to 83% [186]. Samir *et al.*,[185] demonstrated that the median SWE value in the transverse plane was a useful indicator of suspicion of malignancy while other studies indicated that the mean SWE, standard deviation SWE and the longitudinal plane were more useful [187-189]. Most studies concur that there are variable cut-off points and elasticity index parameters that were used in different studies to determine malignancy [30, 31, 190, 191]. Varying points of measurement will often result in conflicting findings regarding the diagnostic performance of SWE. The differences in the findings of these studies can further be attributed to the different vendor machines, user experience, sample size and methodology. The lack of standardized methodology and cut-off points for determining malignancy compromises the reliability of elastography in predicting malignancy in thyroid nodules. Standard cut-off points and statistical parameters for elasticity indices should be established to ensure definitive diagnosis of thyroid malignancy using elastography.

However, it should be noted that elastography has limited utility in cystic, calcified, and multinodular nodules which contributes to unreliable results with using the technique [92, 184, 186]. Nodules with multiple calcifications will naturally have increased stiffness values which may result in increased false-positive rates for thyroid malignancy when examined with elastography.

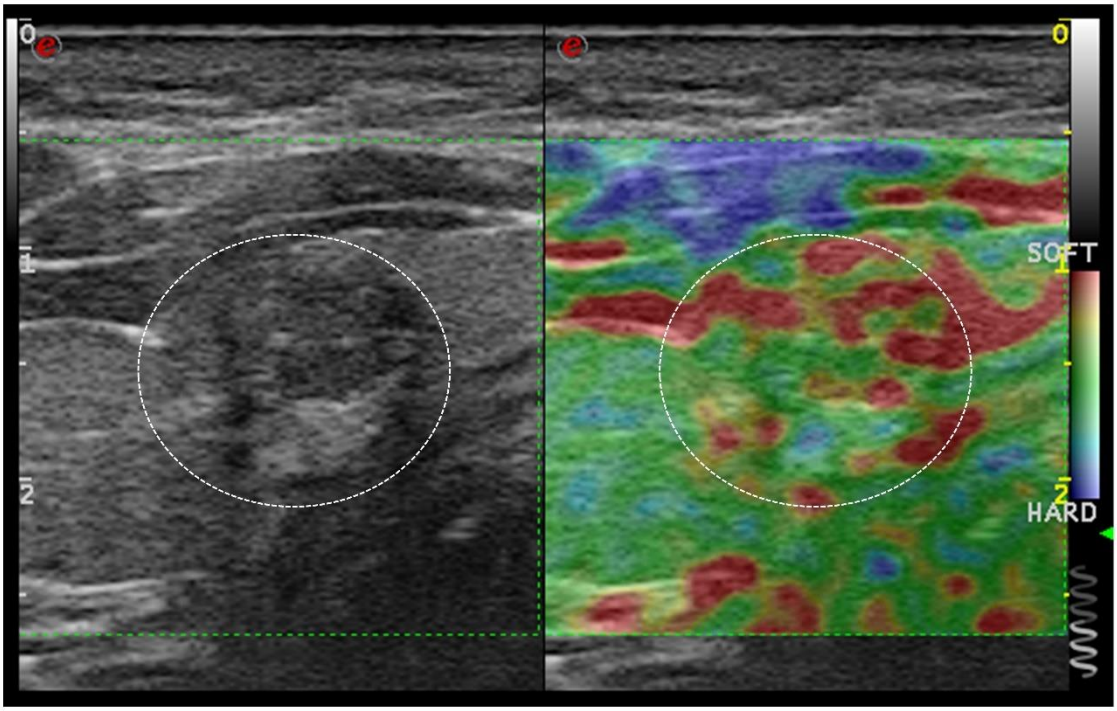


Figure 2:9 Representation of a nodule in grey scale ultrasound (white dotted oval area) and the corresponding SE colour map of the nodule area to indicate stiffness.

The nodule was predominantly soft as indicated by the scale with red representing soft tissue in red and blue representing hard tissue. The 30-year-old male subject had a benign cytology diagnosis.

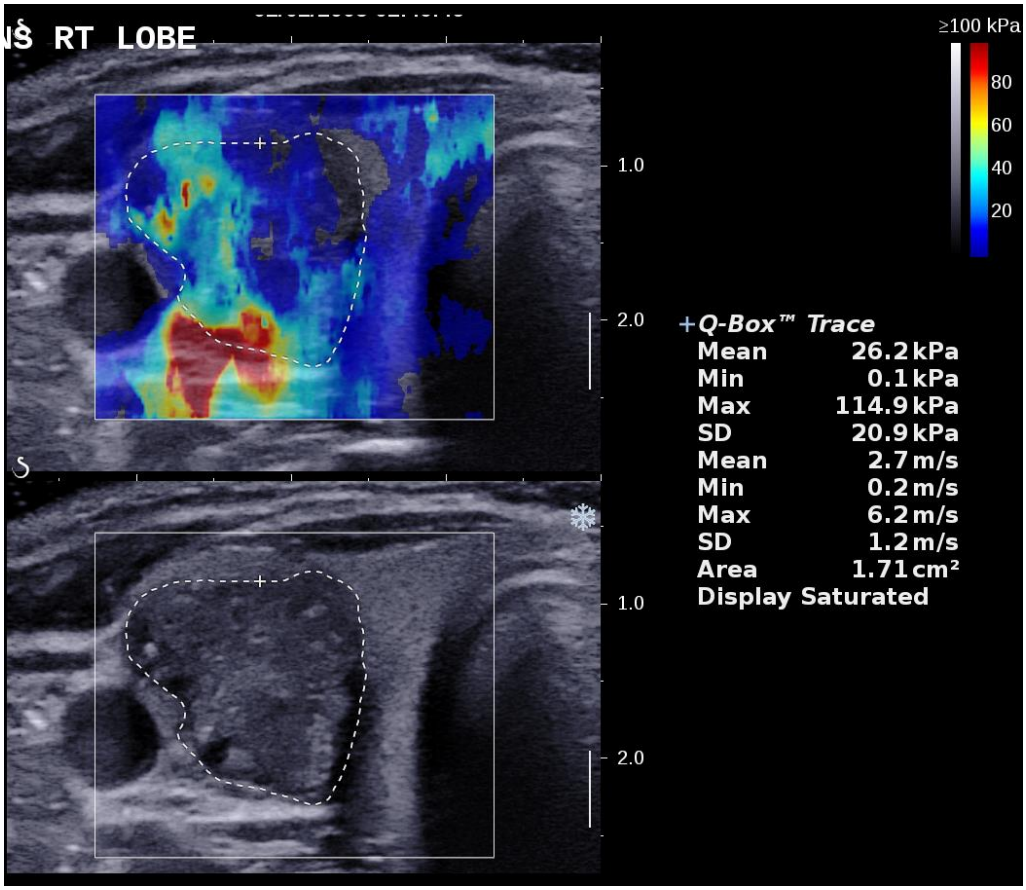


Figure 2:10 A transverse scan of SWE image (above) with the grey-scale view (below) of a malignant thyroid nodule in a 56-year-old female patient.

The maximum elasticity index (Max) of 114.9 kPa was suggestive of thyroid tissue hardness beyond the upper limits of the scale on the right. The histopathology results confirmed papillary thyroid carcinoma.

2.4.1.3.1 Elasticity assessment of nodules with equivocal cytology results

Studies determining the diagnostic utility of SE in nodules with equivocal cytology have yielded confounding results. Using Q-elastography in nodules with equivocal cytology, it was determined that strain ratio had increased diagnostic accuracy for nodules greater than 1cm [192]. This implies that nodule size may influence elastography results in nodules with indeterminate cytology as it does in conventional ultrasound assessment. Rago *et al.,* [193] found SE to have good diagnostic accuracy in characterising thyroid nodules with indeterminate cytology with the potential of aiding in the selection of patients for surgery. However, other studies found no value with SE in the evaluation of thyroid nodules with equivocal cytology as findings had low diagnostic accuracy [194, 195]. The difference in methodology, user experience and sonographic scanning technique was likely the major contributing factors to the conflicting results since SE is a highly subjective and operator-dependent approach.

Presently, there are limited studies that have evaluated the role of SWE in nodules with equivocal cytology. SWE has been suggested to potentially raise the specificity for the pre-surgical selection of nodules with equivocal cytology results, however, more prospective studies are needed to verify this [32, 34]. A recent meta-analysis yielded pooled sensitivities of 83.8% and 71.5% for SWE and SE use in indeterminate cytology, while the pooled specificities were 87.2% and 85.3% respectively [196]. Bardet *et al.,* [34] showed that SWE in thyroid nodules with equivocal cytology had a sensitivity of 85% and a specificity of 94% at a cut-off value of 65kPa. SWE failed to distinguish malignant and benign nodules yielding similar SWE parameters in both benign and malignant nodules, although there were higher elasticity indices for classic PTC variants in their study. Similar studies with the use of SWE in thyroid nodules with equivocal cytology are lacking as well as follow-up studies to correlate prior SWE findings with subsequent final diagnosis at histopathology. The limited study of the value of elastography in thyroid nodules with equivocal cytology reflects the need for further

assessment of objective elastography methods to reduce unnecessary surgical excisions for confirmatory diagnosis.

2.4.1.4 Multi-parametric ultrasound imaging

As previously highlighted, grey scale ultrasound is the primary ultrasound modality for the evaluation of thyroid nodules. However, due to technological advancements, combined techniques involving other ultrasound modalities have been suggested to improve the diagnostic accuracy of grey scale ultrasound alone in thyroid nodule characterization. Multi-modal or combined approaches often involve Doppler ultrasound, elastography techniques and/or CEUS techniques in addition to grey scale ultrasound assessment.

A recent study demonstrated that multi-modal ultrasound thyroid imaging involving grey scale ultrasound, real time elastography (RTE) and SMI resulted in improved sensitivity, specificity and accuracy overall (65% to 94%; 69% to 87% and 67% to 91%, respectively) [178]. Similarly, another study involving CEUS, SMI and grey scale ultrasound, reported that the diagnostic accuracy was highest and improved from an AUC of 0.65 to 0.90 and sensitivity of 40% to 92.5% when SMI was combined with CEUS in the evaluation of moderate suspicion thyroid nodules [197]. Carneiro-Pla [27] indicated that the combination of grey scale, SWE and Doppler ultrasound can help determine which nodules should be biopsied or not. Similarly, for the stratification of intermediate to high-risk suspicion thyroid nodules, Stoian *et al*., [198] found a comparable sensitivity (91.71% vs 96.43%), and improved specificity and accuracy (49.01% vs 80.88% and 57.85% vs 91.18%, respectively) between grey scale ultrasound alone and grey scale ultrasound combined with RTE and 3-D vascularity. Contrarily, a different study found no additional value or significant difference in diagnostic performance from a combination of RTE, SMI and grey scale ultrasound assessment to grey scale ultrasound alone [199].

Multi-parametric ultrasound evaluation has been proposed as a potential solution in the diagnosis of challenging thyroid nodules with equivocal cytology. There are currently few multi-modal ultrasound studies focused on thyroid nodules with equivocal cytology. In one study, the multi-modal evaluation of grey scale ultrasound combined with SE and CEUS had lower sensitivity and comparable specificity than grey scale ultrasound alone (64% vs 79% and 92% vs 100%, respectively) [200]. Contrarily, Gay *et al.*, [35] showed that a combination of molecular testing with SE, SWE and grey scale ultrasound improved the sensitivity from 73.7% to 89.5% while the specificity remained unchanged. Furthermore, in another study the combination of grey scale ultrasound with SE and volumetric vascularity analysis improved the diagnostic sensitivity, specificity and accuracy when compared to grey scale ultrasound alone (57.1% vs 85.7%; 67.4% vs 88.4% and 64.1% vs 90.3%, respectively) [201].

The difference in the study designs of the aforementioned multi-modal thyroid ultrasound studies presents the major challenge in the comparative assessment of the diagnosis value of multi-modal approaches. Furthermore, the approaches involved mostly the qualitative assessment of elasticity, vascularity and/or CEUS which is highly subjective and prone to inter-observer variances. While techniques involving CEUS yield high diagnostic accuracy, the use of CEUS in the routine clinical diagnostic assessment of thyroid nodules is limited by the risk of allergic reactions with the use of contrast agents and cost implications. However, microvascularity imaging has been determined to have a diagnostic performance comparable to that of CEUS and is readily available and safer to use, hence it is more predominant in the few current multi-modal studies. However, there remains a paucity of studies incorporating computer-aided diagnosis in multi-modal approaches. Since computer-aided techniques are expected to be more objective; the evaluation of their value in multi-modal thyroid nodule assessment would help determine routine clinical adoption. The paucity of objective multi-

modal ultrasound studies in literature and conflicting results in current studies warrants the evaluation of the value of multi-modal ultrasound imaging in the diagnosis of thyroid nodules.

2.4.2 Cross-sectional imaging in thyroid cancer

Computed tomography (CT) and magnetic resonance imaging (MRI) play a very minimal role in the primary diagnosis of thyroid cancer and are regarded as advanced cross-sectional imaging. However, advanced cross-sectional imaging is necessary when patients present with symptoms indicative of extrathyroidal extension of tumour causing compression of recurrent laryngeal nerve or oesophagus such as hoarseness, stridor, or dysphagia [69, 81]. Therefore MRI and CT imaging may be used in the pre-operative planning of large thyroid cancers with local invasion of adjacent local structures to determine the management approach in resectable masses [202]. Furthermore, both CT and MRI of the thorax and the neck are best for the assessment of lymph node involvement particularly for lymph nodes in the deep cervical, substernal, and retro-tracheal regions that cannot be visualised on ultrasound [203, 204]. The primary role of MRI and CT in thyroid cancer evaluation is in the evaluation of disease extent in persistent and recurrent thyroid cancer [44, 81]. Thyroid nodules are often an incidental finding during MRI and CT of the head and neck regions and account for about 20% of thyroid cancer cases [205].

2.4.2.1 CT imaging

CT imaging of the neck often requires the use of iodinated contrast to enable visualization of blood vessels and adjacent structures [206]. In patients with differentiated thyroid cancer, the use of iodinated contrast is contra-indicated in order to avoid interference with subsequent radioiodine scanning and therapy [38, 125]. However, in patients with bulky cervical lymphadenopathy and local invasion of the primary tumour, iodinated contrast can be used in neck CT imaging based on a risk-benefit analysis of additional anatomical assessment

outweighing the risk of delayed RAI therapy [77]. While non-contrast-enhanced CT of the neck is an option, it has poor sensitivity and specificity in characterising tumour composition due to its failure to adequately distinguish cysts and solid lesions and it has poor delineation of multiple nodules [207]. However, a non-contrast CT chest examination is most helpful in the evaluation of mediastinal and lung metastases in the follow-up of follicular carcinoma which tends to have haematogenous spread [38, 208].

Several benefits of CT use in neck imaging include quick scan times, wide availability and easy generation of reformatted images in multiples planes such as axial, coronal and sagittal [209]. The major advantages in thyroid cancer imaging are the ability to evaluate extra-thyroidal invasion of local structures such as the oesophagus, trachea and major blood vessels and the detection of calcifications as evidenced in papillary thyroid carcinoma [44, 115, 210]. CT imaging is best for the assessment of pulmonary metastases and distant metastases [125]. Additionally, substernal masses which are unreachable by routine ultrasound may be biopsied under CT guidance. CT imaging enables cervical lymph node assessment in patients with established carcinoma but equivocal physical examination findings [10]. Calcifications, cystic areas or haemorrhagic components of nodal masses appear more vividly on CT due to increased enhancement [211]. CT best demonstrates calcifications in medullary carcinoma which are usually denser than those seen in papillary carcinoma [125].

The major limitation of CT imaging in thyroid cancer imaging includes the exposure to ionizing radiation compared to ultrasound and MRI, its non-ionising radiation counterparts. Additionally, is the previously stated contra-indication of iodinated contrast use in many thyroid cancer patients who have to undergo subsequent RAI therapy. Other limitations include the failure to distinguish metastatic lymph nodes from small normal lymph nodes and post-surgical changes from recurrent cancer [212]. Due to some of these limitations, CT is often

reserved for follow-up evaluation in metastatic and aggressive disease and is not used in early thyroid cancer diagnosis.

2.4.2.2 MR Imaging

MRI is the preferred cross-sectional imaging modality for the assessment of the local extent of the tumour. MRI is used in demonstrating any tracheal and oesophageal invasion and assessing recurrent adenopathy in thyroid cancer patients [38]. MRI has been determined to have high precision and sensitivity (80-95%) in the assessment of the extent of tumour invasions such as oesophageal involvement, tracheal cartilage invasion and recurrent laryngeal nerve involvement [206]. The area of invasion as evidenced by the degree of contact of adjacent structures and the sole extracapsular invasion appears ill-defined, while metastases demonstrate as cystic foci or enlargement on MRI [125]. Imaging features that are highly suspicious for malignancy and metastases appear to include marked enhancement relative to non-diseased thyroid tissue, haemorrhagic and necrotic appearance and hyperintensity on T2 weighted images [204]. An infiltrative pattern on MRI is often indicative of ATC. MRI is highly precise in detecting skeletal metastases in MTC and best demonstrates rare sites of metastases from thyroid cancer such as the brain and breast than other imaging modalities [203, 206]. Furthermore, MRI is highly accurate for the multi-focal and multi-centric detection of disease and quantification of lobular cancer hence it is the primary imaging modality in evaluating liver metastases [208, 209]. Current MRI machines that can produce small slice thicknesses ensure the precise detection of recurrent lesions and lymph nodes as small as 4 - 5mm in diameter in patients with abnormal I-123 uptake or high Tg levels in the absence of any other suspicious findings [206].

The main advantages of MRI in thyroid cancer imaging are that it is non-invasive, does not involve ionising radiation, and has high soft tissue resolution. Other advantages include the

acquisition of images in the various anatomical planes without the need for reformations and minimal contrast reactions due to the use of an inert contrast agent, gadolinium [209]. However, MRI has the limitation of longer imaging time and higher cost compared to CT. Additionally, due to safety concerns, patients with implanted metallic devices such as pacemakers, and some implanted cochlear devices and prostheses are precluded from an MRI examination. These devices must be thoroughly certified to be MRI-compatible to exclude any possibility of interaction with the magnetic field in the MRI machine. Therefore, thorough patient screening for any metallic devices is required before an MRI examination. Another concern with MRI is claustrophobia, which results in some patients being given anti-anxiety medication before the examination. Although open-bore MRI units have become available, their limitation is lower resolution than high field closed units. This limitation impedes the thyroid imaging of very small (2-3mm) lymph nodes which cannot be detected with open MRI units [206].

MRI and CT both lack specific imaging appearances for characterization of lymph nodes as malignant and also have poor sensitivity for detection of nodal involvement in thyroid cancer patients in comparison to ultrasound [125]. However, the use of these cross-sectional imaging modalities is most helpful in guiding surgical planning as it helps in ascertaining the extent of the disease.

2.4.3 Nuclear medicine imaging

2.4.3.1 Radioactive iodine scintigraphy

Radioactive imaging with iodine and Technetium (^{99m}Tc) as preferred first-line diagnosis of thyroid cancer has declined due to the emergence of more specific imaging modalities [208, 213]. Scintigraphy is now typically performed for post-surgery follow-up and assessment of recurrent and metastatic diseases. Radioiodine scintigrams are usually positive because papillary thyroid cancer has follicular cell origin and hence has an affinity for iodine [206]. Iodine-131(^{131}I), iodine-123 (^{123}I) and technetium pertechnetate ($^{99m}TcO_4$) are the common

radioisotopes used in RAI studies for thyroid cancer. Of these three common radioisotopes, ^{131}I has the most radiation-absorbed dose in normal thyroid tissue, hence it is usually used as an option when ^{123}I and $^{99m}TcO_4$ are not available [214]. However, iodine radioisotopes are usually preferred due to high false negatives with technetium pertechnetate [205]. The radioisotopes are used in thyroid scintigraphy to determine the functionality of a thyroid nodule, whether it is hyper-functioning ("hot"), hypo or non-functioning("cold") or iso-functioning ("warm"), when serum thyrotropin (TSH) levels are below normal levels or in nodules with equivocal FNAC results [210, 215].

"Cold" nodules have a reduced iodine uptake and are usually indicative of a malignant disease process, while "hot" nodules have increased regulation of TSH and tend to be linked to benign hyperthyroidism. However, some studies have suggested that an average of about 15-21% of "cold" nodules are malignant, while very few "hot" nodules may also exhibit malignancy particularly in the unlikely occurrence of a carcinoma co-existing within a hyper-functioning nodule [44, 214]. Due to this lack of specificity and limited information, radionuclide scans are considered unreliable in characterizing benign and malignant thyroid lesions and have limited use in the routine evaluation of thyroid nodules [38]. The bone scan and ^{18}F-fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET) are essential complementary imaging modalities in suspected extensive metastatic disease when CT and MRI studies of the neck, thorax, and abdomen yield inconclusive results in detecting metastases [208]. Therefore, bone scintigraphy can be used as the primary imaging modality for the assessment of bone metastases in such cases. However, nuclear medicine studies have poor accuracy in detecting lymph node metastasis when compared to CT, MRI and ultrasound. Regardless of nodule size or appearance, abnormal uptake by iodine-123 or a growing size with PET imaging are major predictors for lymph node malignancy [206].

2.4.3.2 Positron emission tomography (PET) imaging

The role of PET imaging in thyroid cancer is usually staging, follow-up and assessment of patient prognosis. It is not normally used for primary diagnosis as differentiated thyroid cancers can produce variable results in PET [213]. However, PET plays a significant diagnostic role in cases with negative whole body RAI scans but having elevated thyroglobulin levels.^{18}F-fluorodeoxyglucose (^{18}F-FDG) is the most commonly used radiopharmaceutical in PET imaging because of its active uptake and retention by cancerous cells [216].^{18}F-FDG-PET is important in follow-up cases to detect recurring or metastatic disease especially in high-risk patients and cases of elevated TSH-stimulated thyroglobulin [215]. After ultrasound, ^{18}F-FDG-PET is highly reliable in detecting lymph node metastases [217].

The increased utility of CT, MRI and PET scanning in thyroid imaging has resulted in an increasing number of incidental findings of thyroid nodules with 16% being attributed to CT and MRI, while 2% are from PET scans [202]. Incidentalomas found on ^{18}F-FDG-PET images provide physiological and tumour location information and therefore suggest malignancy more than incidental findings on ultrasound and CT imaging [218]. Some studies have suggested that the malignancy rate in F-FDG-PET incidentalomas can be up to 40% while the sensitivity can be almost 90% [219, 220]. ^{18}F-FDG-PET has the limitation of poor spatial resolution and lack of anatomical detail in PET images and therefore results in the lack of accurate localization of FDG-positive lesions [212, 215]. Furthermore, there is the challenge of uptake of ^{18}F-FDG-PET by normal muscle and lymphoid tissues and the misinterpretation of autoimmune thyroid disease as malignancy [212, 221]. The variable sensitivity of ^{18}F-FDG-PET precludes its use in the initial assessment of metastatic disease although the sensitivity improves with elevated calcitonin levels [203]. However, ^{18}F-FDG-PET-CT has better lesion and uptake characterisation than ^{18}F-FDG-PET [218]. This alludes to the additional benefit of cross-sectional anatomical visualisation with CT in combination with the physiological information already provided by ^{18}F-FDG-PET.

2.4.3.3 PET-CT

The emergence of hybrid PET/CT scanners has raised the diagnostic accuracy in restaging DTC patients and guiding favourable treatment options and reduced discrepant findings from the sole use of PET or CT [215, 217]. PET-CT imaging is considered highly sensitive and specific in the early detection of thyroid cancer. In a study by Zoller *et al*., [217], PET-CT altered the sole PET diagnoses in 77% of the patients while the treatment plan was altered in 23% of the patients. In other studies with iodine-negative but thyroglobulin positive cases, PET-CT altered clinical management in 23% to 51% of patients resulting in a change of treatment in 48% PET-positive patients [222]. Another study found that ^{18}F-FDG-PET/CT has excellent sensitivity of about 88.6% and specificity of about 89.3% in the detection of metastatic disease in DTC patients with elevated or rising Tg levels [223]. In some studies of thyroid nodules with equivocal cytology, ^{18}F-FDG-PET/CT demonstrated a sensitivity and negative predictive value of up 100% [224]. The fusion of PET and CT, therefore, results in precise anatomical information that results in increased sensitivity, specificity and diagnostic accuracy and influences therapeutic approach. Combined PET-CT has the advantage of a shorter scan time of about 2 minutes thereby resulting in not only anatomical information but also a faster throughput of patients in comparison with PET which takes about 30 minutes [216]. The hybrid technique additionally results in better computational attenuation correction of PET photons within the body [225].

However, ^{18}F-FDG-PET-CT has the limitation of lacking specificity as ^{18}F-FDG uptake may not be specific for thyroid cancer and false-positive uptake can occur due to other malignant cells or in active muscle cells [216]. Furthermore, due to variable FDG-uptake in MTC, PET-CT is unreliable in the evaluation of persistent disease in patients with this type of thyroid cancer [211]. Ho *et al*., [221] found a 3.7 % prevalence of false-positive uptake on ^{18}F-FDG PET-CT of patients with no prior history of thyroid malignancy and of those verified by cytology and histopathology, only 14% had thyroid malignancies. Image interpretation that includes

quantitative measurement of glucose standard uptake values (SUV), ^{18}F-FDG uptake and CT attenuation may improve the accuracy of PET-CT in focal thyroid lesion characterisation although focal ^{18}F-FDG-PET-CT incidentalomas, often have a high suspicion of malignancy [226]. In one study, suspicious ultrasound findings in combination with an elevated SUV of about 5.3, yielded a sensitivity of 82% in differentiating benign and malignant thyroid nodules [227]. Nonetheless, PET-CT is a relatively expensive and not easily accessible procedure.

2.5 Computer-aided diagnosis in thyroid ultrasound imaging

Computer-aided diagnosis (CAD) systems have emerged over the past years as non-invasive techniques to complement radiologists' interpretation. CAD methods are considered a solution to potentially overcome subjective interpretation limitations. CAD methods often rely on machine learning (ML) techniques and quantitative approaches that use statistical and data mining algorithms to characterise predictive ultrasound features. In the quantitative analysis of textural grey scale sonographic features clinical and non-clinical features are extracted from sonographic images using data mining and statistical techniques [228, 229]. The typical clinical features include nodule composition, shape, calcification, margins and echotexture while non-clinical features include the regularity of pixels, coarseness and smoothness of image [230]. The quantitative analysis of Doppler ultrasound features often involves the regional segmentation of thyroid nodule vascularity into peripheral and central regions to determine malignancy [231]. Due to the computational data training and validation techniques of CAD algorithms, CAD methods are considered more objective, reproducible and accurate in ultrasound imaging diagnosis and reporting [232]. This, in turn, results in standardized predictive ultrasound features for benign and malignant thyroid nodules thereby limiting potential human biases.

2.5.1 Grey scale CAD in thyroid nodule assessment

CAD software can be embedded within the ultrasound unit or be used as an isolated program for offline image analysis. Grey scale ultrasound CAD software is equipped with selected TIRADS for diagnostic purposes.

Presently, there are two globally-approved commercial thyroid ultrasound CAD software, AmCAD-UT (AmCad Biomed, Taipei, Taiwan) and S-Detect (Samsung Medison Co. Ltd., Seoul, South Korea). A few studies have investigated the diagnostic performance of these thyroid ultrasound CAD methods for potential application in the clinical diagnostic workflow. While it has been suggested that CAD approaches based on deep-learning (DL) methods have the potential to outperform even radiologist experts in diagnostic accuracy of thyroid ultrasound, recent studies have shown comparable diagnostic performance [233-236]. The comparison between studies based on classical ML and DL-based CAD also showed comparable pooled sensitivity (86% vs 89%), specificity (85% vs 84%) and diagnostic odds ratio (37.41 vs 40.87%) in a recent meta-analysis [237]. The current commercially available CAD software is based on both ML and DL methods. The lack of sufficient evidence for the additional value of the CAD software in comparison with experienced clinicians limits the routine clinical adoption of the techniques. As various researchers and clinicians are developing thyroid ultrasound CAD systems globally, there are likely to be multi-factorial influences on CAD diagnostic performance.

The diagnostic value of CAD systems for grey scale ultrasound of thyroid nodules has been determined in several studies. The diagnostic performance outcomes of some of the studies were summarized in a systematic review that our group conducted [238]. Table 2.1. shows the diagnostic performance outcomes from the review and those from more recent studies. Some studies have shown that the sensitivity (SEN) and negative predictive value (NPV) of CAD are comparable to that of radiologists while the specificity tends to be reduced [239-246]. Contrarily, a

few studies reported a higher specificity but lower sensitivity with CAD than with subjective interpreters [247-249]. In other studies, the diagnostic performance of radiologists, especially those less experienced, was improved with the assistance of the CAD, resulting in further improved diagnostic sensitivity or specificity and overall accuracy [236, 241, 250, 251]. Contrarily, a recent multi-centre and multi-reader study showed that the use of CAD improved the overall diagnostic accuracy of the users regardless of experience, and junior readers with CAD outperformed senior readers without CAD [252].

Few studies have evaluated the diagnostic performance of CAD based on multiple TIRADS. Table 2.1 summarises some of the studies whereby the diagnostic performance of either the CAD or radiologists or both were based on specific TIRADS. Furthermore, the AmCAD CAD software is currently the only commercial thyroid CAD software that has multiple TIRADS embedded in it, while S-Detect only has K-TIRADS (KTA/KSThR) and ATA. ATA and K-TIRADS using S-Detect CAD have been recently indicated to have comparable sensitivity with radiologists using the same set of TIRADS, however, the specificity is still lower [245]. There are currently few studies involving the assessment of CAD performance based on multiple TIRADS using AmCAD. Reverter *et al.*, [242] concluded that AmCAD CAD analysis based on ATA, EU and AACE in comparison with radiologist subjective assessment using ATA resulted in comparable sensitivity for ATA (87%), whereas EU and AACE had slightly lower sensitivity than the radiologist (85.2% and 81.5% respectively). Subjective assessment by the radiologist yielded a significantly higher specificity of 91.2% in comparison with the highest specificity yielded with CAD (68.8%) achieved with ATA TIRADS. Contrarily, a recent multi-centre study using AmCAD found that CAD using KSThR resulted in the highest sensitivity (90.5%) and accuracy (0.75) than using all other TIRADS supported by the software [246]. However, the CAD-KSThR specificity was lower (49.63%) than radiologists of different levels of experience (75%-90%) using the same TIRADS for subjective assessment. The

aforementioned studies had different methods of selecting the universal TIRADS for comparing the diagnostic performance of CAD with that of subjective assessors. The paucity of similar studies means that the choice of the best TIRADS to use with CAD remains vague and thus necessitates further exploration of the influence of different TIRADS on CAD performance.

Table 2.1: Summary of diagnostic performance assessment outcomes from different studies comparing grey scale ultrasound thyroid CAD with radiologists using different criteria

Reference	CAD	Diagnosis criteria	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	AUROC
Choi *et al.*, [239]	S-Detect	CAD	90.7	74.6	72.2	91.7	.83
		Rad_K-TIRADS	88.4	94.9	92.7	91.8	.92
Gao *et al.*, [240]	Non-commercial	CAD	96.7	48.5	81.3	86.2	.73
		Rad_Kwak-TIRADS	96.2	75.7	90.2	89.7	.87
		Rad_ATA	95.4	78.6	91.2	88.0	.83
		Rad_ACR	90.0	76.7	90.0	76.7	.86
Gitto *et al.*, [247]	S-Detect	CAD	21.4	81.3	25.0	78.0	ND
		Rad_K-TIRADS	78.6	66.7	40.7	91.4	ND
Yoo *et al.*, [241]	S-Detect	CAD	80.0	88.1	83.3	85.5	.84
		Rad_ K-TIRADS	84.0	95.5	93.3	88.9	.90
		Rad + CAD	92.0	85.1	82.1	93.4	.89
Wang *et al.*, [235]	Non-commercial	CAD	90.5	89.9	95.2	81.0	.90
		Rad_French-TIRADS	93.8	78.0	90.4	85.0	.86
Jeong *et al.*, [251]	S-Detect	Expert Rad_ K-TIRADS	84.1	96.4	94.9	88.5	ND
		Expert + CAD	88.6	83.9	81.3	90.4	.86
		User 1 + CAD	70.5	80.4	73.8	77.6	.75
		User 2 + CAD	75.0	73.2	68.8	78.8	.74
		User 3 + CAD	70.5	73.2	67.4	75.0	.72
Reverter *et al.*, [242]	AmCAD	Expert Rad_ATA	87.0	91.2	90.5	90.9	.88
		CAD_ATA	87.0	68.8	64.5	86.3	.72
		CAD_EU	85.2	50.2	50.1	82.6	.71
		CAD_AACE	81.5	53.2	51.8	80.8	.70
Xia *et al.*, [243]	S-Detect	Expert Rad	81.1	83.5	84.6	79.8	.82
		CAD-dichotomous	90.5	41.2	63.2	79.5	.66

Kim *et al.*, [244]	S-Detect	S-Detect 1	80.2	82.6	75.0	86.3	.81
		S-Detect 2	81.4	68.2	62.5	84.9	.75
		Rad_ K-TIRADS	84.9	96.2	93.6	90.7	.91
		Rad + S-Detect 1	91.9	81.1	76.0	92.2	.87
		Rad + S-Detect 2	93.0	67.4	65.0	93.7	.80
Jin *et al.*, [236]	Non-commercial	CAD_ACR	80.6	80.1	76.0	84.1	.88
		Junior Rad_ACR	87.5	57.9	61.8	85.7	.73
		Junior Rad + CAD	78.1	80.0	73.1	84.0	.83
		Intermediate Rad_ACR	95.1	63.9	67.3	94.3	.80
		Senior Rad_ACR	87.4	83.2	88.0	83.4	.91
Li *et al.*, [248]	AmCAD	CAD	76.9	87.5	86.9	77.8	.82
		Junior Rad	82.6	70.4	70.4	82.6	.76
		Senior Rad	86.9	77.8	76.9	87.5	.82
Fresilli *et al.*, [249]	S-Detect	CAD_K-TIRADS	70.4	87.5	65.5	89.7	.79
		Student_K-TIRADS	70.4	76.3	50.0	88.4	.73
		Resident_K-TIRADS	74.1	85.0	62.5	90.7	.80
		Expert_K-TIRADS	81.5	88.8	71.0	93.4	.85
Han *et al.*, [245]	S-Detect	CAD_K-TIRADS	89.2	50.2	51.0	88.9	ND
		Rad_K-TIRADS	91.6	76.7	69.5	94.0	ND
		CAD_ATA	87.4	63.4	58.2	89.7	ND
		Rad_ATA	83.8	80.8	72.3	88.3	ND
Ye *et al.*, [246]	AmCAD	CAD_KSThR	90.5	49.6	66.3	82.7	.75
		Junior Rad_KSThR	72.2	75.6	76.3	82.0	.74
		Junior Rad + CAD	89.9	73.3	78.6	86.8	.82
		Intermediate Rad_KSThR	82.7	85.9	86.5	85.6	.82
		Expert Rad_KSThR	88.5	90.4	90.9	87.8	.89

SEN, sensitivity; SPEC, specificity; PPV, positive predictive value; NPV, negative predictive value; AUROC, area under ROC curve; Rad, radiologist; ND, not determined; CAD – computer-aided diagnosis; CAD-dichotomous, possibly benign/possibly malignant

2.5.2 Doppler ultrasound CAD in thyroid nodule assessment

A limited number of studies have evaluated thyroid nodule vascularity using CAD methods. The Doppler ultrasound assessment of thyroid nodular vascularity in most current studies has largely been based on subjective grading qualitative methods. Doppler ultrasound CAD is often based on stipulated cut-off points for vascularity indices (VI) used in differentiating peripheral and central vascularity in thyroid nodules. The comparison of qualitative CDUS vascularity analysis with quantitative central vascular area analysis showed higher sensitivity (90% vs 67.5%) with quantitative assessment than qualitative vascular assessment in one study [253]. Similarly, Ying *et al.,* [231] developed a CAD algorithm for the objective semi-quantification of thyroid parenchyma vascularity in Doppler sonograms. This same algorithm demonstrated the objectivity and accuracy of CAD quantification of intranodular vascularity in the distinguishing reactive and metastatic lymph nodes [168, 169]. Another different study showed that thyroid nodular regional VIs based on CDUS significantly increased diagnostic accuracy from 58.6% to 79.3% when combined with grey scale ultrasound features for thyroid nodule differentiation [23]. Therefore, this suggests that combined ultrasound techniques involving quantitative vascularity assessment may potentially result in improved diagnostic performance.

However, a different study that computed VIs for central and peripheral vascularity densities from PDUS images suggested that even with high sensitivity (84.8%), intranodular vascularity was not a reliable predictor of malignancy as benign nodules had higher vascularity VIs in both central and peripheral regions [254]. Furthermore, another study suggested that the size of the thyroid nodule affects the quantitative vascularity assessment based on normalised VIs with higher specificity (100% vs 54.3%) being achieved for nodules < 2cm than for all nodules [255].

Although these few studies largely suggested that quantifying intranodular vascularity using automated regional segmentation is more objective in vascularity assessment for thyroid nodule

differentiation, the differences in their methodology present clinical adoption challenges. Thus, the value of quantified VIs requires further exploration to determine the acceptable thresholds for optimised sensitivity and specificity. Furthermore, since the diagnostic value of combined ultrasound assessment involving CAD approaches in thyroid nodule diagnosis is scanty, this necessitates the evaluation of multi-modal approaches involving CAD.

2.6 Basis for the study

The global increase in the incidence of thyroid cancer has resulted in concerns about over-diagnosis due to the advancement in diagnostic imaging technologies and the overuse of FNAC. Overdiagnosis results in unnecessary surgical procedures particularly for non-lethal indolent micro-carcinomas. The surgical treatment of thyroid cancers has surgical risks such as facial palsy that can arise from potential damage to the recurrent laryngeal nerve. Furthermore, there cost and lifestyle implications post-surgery such as lifetime thyroid hormone replacement therapy.

FNAC is the minimally-invasive pre-surgical reference standard for the diagnosis of thyroid nodules. However, it has the challenge of the equivocal cytology nodules category which often undergo surgery for definitive diagnosis, yet most nodules in this category tend to benign. Ultrasonography is the primary non-invasive modality used in the diagnosis of thyroid nodules. While it is available in multiple imaging modes, grey scale ultrasound features are primarily used to differentiate benign and malignant nodules and are the basis for most of the current TIRADS. However, no sole grey scale ultrasound is highly predictive of malignancy on its own and the reliability of grey scale ultrasound features in the diagnosis of thyroid nodules with equivocal cytology remains unclear. The assessment of grey scale ultrasound features in combination with elasticity and vascularity imaging techniques is purported to increase the diagnostic accuracy although conflicting findings in the literature persist due to diverse methodologies.

Advanced imaging modes of ultrasound such as SWE for quantitative tissue elasticity assessment have the potential of improving the diagnosis of thyroid nodules, particularly those with equivocal cytology results. However, presently few studies have evaluated this and SWE cut-off criteria and optimal measurement parameters for thyroid nodule diagnosis remain

unclear and unstandardized. Although advanced ultrasound imaging modes of thyroid nodule vascularity assessment have emerged, there remains a lack of the evaluation of the diagnostic value of quantitative assessment in. More so, the diagnostic accuracy of novel microvascularity techniques such as AngioPLUS in thyroid nodule differentiation lacks evaluation both qualitatively and quantitatively. While thyroid ultrasound CAD techniques have been suggested to be more objective and potentially more accurate than subjective assessors, few studies suggest otherwise. However, there is a lack of studies evaluating the diagnostic accuracy of CAD techniques in combination with other ultrasound imaging modes. Hence the value of multi-modal ultrasound imaging in thyroid nodule differentiation remains ambiguous yet it may have the potential of improving diagnostic accuracy.

The ethical board approved a sample size of 150 patients for prospective assessment in the current study. This was based on the specificity sample size calculation using Buderer's formula assuming a 10% drop-out rate and 40% prevalence rate. The present study investigated the diagnostic accuracy of grey scale ultrasound CAD in comparison with subjective ultrasound assessment using different TIRADS. In addition, the diagnostic performance of AngioPLUS based on the qualitative and quantitative regional vascularity assessment was determined in vascular nodules. Lastly, the value of SWE in determinate and equivocal thyroid nodules in combination with EU TIRADS was evaluated, and the assessment of a multi-modal approach involving grey scale ultrasound, AngioPLUS and SWE was conducted for equivocal thyroid nodules. The results of the present study are anticipated to aid in the evidence-based application of individual advanced ultrasound techniques and the multi-modal use of ultrasound in thyroid cancer diagnosis thereby optimising ultrasound imaging use and improving thyroid nodule diagnosis.

Chapter 3

Study One: Diagnostic Performance Evaluation of Computer-Aided Diagnosis and Human Assessment of Grey Scale Ultrasound Features of Thyroid Nodules based on TIRADS

3.1 Introduction

Various RSS or TIRADS were developed to aid the stratification of risk of malignancy of thyroid nodules based on several suggestive sonographic features. TIRADS are purported to improve consistency in subjective interpretation and limit inter-observer variabilities [71, 135, 256, 257]. Nevertheless, divergences amongst different TIRADS still exist due to varying malignancy risk estimation criteria for suspicious sonographic features [133]. Therefore, based on the diverse malignancy prediction criteria, the use of TIRADS can be subjective and dependent on the clinical approach of the clinician [2]. Hence, currently, there is no consensus on the universal standard for the best RSS to use in thyroid nodule differentiation.

An online-based multiple RSS malignancy risk scoring calculator based on subjective interpretation of sonographic features was developed in the past recent years [258]. The RSS outputs available with this online risk calculator are the AACE/ACE/AME, ATA, KSThR, and French TIRADS (now updated and known as EU TIRADS). This predictive model has been evaluated in comparison with other similar subjective interpretation-based models and found to be highly accurate and reliable in thyroid nodule differentiation [259]. On the other hand, with the evolution of artificial intelligence (AI), computer-aided diagnosis (CAD) systems have been proposed as a more objective and consistent method of thyroid nodule differentiation than human visual assessment [229, 260]. Recent studies have shown that thyroid CAD systems have a

diagnostic performance that is comparable to that of experienced radiologists with combined techniques having more potential for superior performance [234, 261]. The current globally-approved commercial thyroid CAD software with multiple TIRADS computations is AmCAD-UT (AmCad Biomed, Taipei, Taiwan). This CAD software has been previously evaluated in comparison with human interpreters in a few studies. Some studies have evaluated its diagnostic performance in comparison with clinical experts and radiologists [242, 246] and its role in guiding sonographers in diagnosing space-occupying thyroid lesions [262]. Although AmCAD-UT settings can be adjusted for optimised diagnostic performance, the previous studies only assessed the diagnostic performance using the default settings with a limited comparison of multiple TIRADS.

Although the online risk calculator (based on a prediction model) and the CAD software mentioned above both offer automatic multiple-TIRADS output, currently no study has comparatively evaluated their diagnostic performance for paired multiple TIRADS for consideration of clinical adoption. Furthermore, plausible evidence for the additional value of CAD from the studies on different non-commercialized thyroid CAD technologies remains vague due to the variable results for the different TIRADS [233].

The present study began with the evaluation of the diagnostic performance of AmCAD-UT at varied detection sensitivity settings of different ultrasound features for thyroid nodule differentiation based on six different TIRADS within the software. This helped validate the best detection sensitivity setting and TIRADS for the subsequent analysis and comparisons with human visual assessments. Once the best setting was established, then the diagnostic performance metrics of computer-assisted subjective analysis using the online risk calculator and AmCAD-UT based on the same multiple TIRADS were analysed and compared. The findings of this study could potentially assist in determining a diagnostically efficient approach and TIRADS considerations for clinical diagnostic workflow adoption.

The PhD candidate has published the results of the CAD software validation and the comparative analysis of CAD and computer-assisted subjective interpretation that are presented in this chapter [263, 264].

3.2 Validation of software sensitivity detection settings

3.2.1 Materials and methods

3.2.1.1 Study type

This analytical observational retrospective study was approved by the Human Research Ethics sub-committee at the Hong Kong Polytechnic University before the study was conducted. A consecutive case analysis approach was used for the data collection of thyroid nodule ultrasound images. Informed consent was waived for this retrospective study.

3.2.1.2 Data sources

Thyroid nodule images were retrieved from image archives of thyroid ultrasound studies previously conducted by our research group and from an open access thyroid ultrasound image database, Digital Database of Thyroid Ultrasound Images (DDTI) (Universidad Nacional de Colombia, CIM@LAB and Instituto de Diagnostico Medico (IDIME), Bogota, Colombia) [265]. A total of 205 images from 198 patients were retrieved from both thyroid ultrasound image sources with 104 images being from studies conducted by our research group between February 2013 and December 2014. Images from the previous studies by our research group were all acquired using a Supersonic Aixplorer ultrasound machine (SuperSonic Imagine, Aix-en-Provence, France) in conjunction with a 4-15 MHz linear transducer [23, 190]. This was the first time the images were used in a grey-scale ultrasound computer-aided diagnosis analysis study. The 101 images retrieved from the online database were obtained using the following types of

ultrasound machines: a TOSHIBA Nemio 30 and a TOSHIBA Nemio MX (Canon Medical Systems, Tochigi, Japan) in conjunction with 12 MHz linear and convex transducers [265].

3.2.1.3 Image selection criteria

The inclusion criteria for ultrasound image selection were diagnostically acceptable thyroid nodule B-mode ultrasound images from adult patients (≥18 years old) who had been evaluated with ultrasound imaging for thyroid cancer suspicion and had final cytological and/or histopathological diagnosis results. Two thyroid surgeons with extensive experience had conducted the fine-needle aspiration cytology of thyroid nodules and provided the cytological and/or histopathological results. Images from the DDTI database had a cytological diagnosis as had been determined by experts [265]. The exclusion criteria were ultrasound images with indistinct nodules, unclear boundaries or margins and incomplete cytology or histopathology.

Areas clearly demonstrating the nodule were selected and separated from the entire image and the new nodule-specific images were saved in JPEG format using codes.

3.2.1.4 CAD analysis of the thyroid nodule images

An independent rater (the PhD candidate - N.C) who had 2-years' thyroid ultrasound experience at the time performed the CAD analysis using the AmCAD-UT thyroid CAD software after a month of training in using the software. The user was blinded to the cytology and/or histopathology results.

3.2.1.4.1 CAD ROI-selection

The coded JPEG images were uploaded onto the AmCAD-UT software user interface for analysis. The AmCAD-UT software has 3 methods for selecting the region of interest (ROI): 1.) manual outlining; 2.) the semi-automated method of selecting four contour points on the nodule which the software then uses for detecting the nodule; and 3.) the automated nodule

recognition method. For this study, the ROI was manually outlined from the thyroid images based on the visualized nodule margins to ensure a standardized uniform approach for all nodules as both the automated and semi-automated methods were found to sometimes miss some nodule areas during the training period. After selecting the ROI, the user then selected and adjusted the different settings for ultrasound feature analysis before confirming the analysis for the diagnosis output. Outlining the nodule and the software computation took less than 2 minutes overall for each nodule. The final diagnosis from the CAD system was documented based on the criteria of the different risk stratification systems within the software.

3.2.1.4.2 CAD settings selection

The AmCAD-UT software can be adjusted for detection sensitivity within pre-determined ranges for margins (1 to 5), hyperechoic foci (2.0 to 4.0) and anechoic areas (0 to 0.5), while it allows flexible control of visualization enhancement for echogenicity (-50 to 50) and texture (10 to 100) analyses. The detection sensitivity increases as the settings for the different ultrasound features are increased except for the hyperechoic foci setting which has an inverse relationship. At its development phase, the standalone diagnostic performance of AmCAD-UT established that the ultrasound feature "hyperechoic foci" is affected by and dependent on detected "anechoic areas". The highest diagnostic performance was determined to be over 90% at a setting of 3.5 for the different ranges of "anechoic areas" setting [266]. The "anechoic areas" settings 0.5 and 0.2 yielded similar high diagnostic performance whereas for margins the best performance was with the 2.0 and 3.0 for the 3 different ultrasound machines used in that testing phase. At the commonly used default setting median values are used for all the parameter settings. Based on this background, this current study sought to determine the setting for optimised diagnostic performance between the default settings and the "hyperechoic foci" maintained at 3.5 with variations of "anechoic areas" and "margins" at settings that previously achieved the highest diagnostic performance during the software development phase testing.

"Echogenicity" and "texture" parameter settings were consistently maintained to those of the default setting (0 and 33 respectively) for the comparative analyses. These two parameter settings mainly influenced subjective visualization of the images without a change in CAD diagnosis output during our pilot testing of the software. The different settings used for the comparisons are tabulated in Table 3.1.

Table 3.1: Different AmCAD-UT settings adjustment for comparative analysis of diagnostic performance

Name of setting	Anechoic area	Hyperechoic foci	Margin	Echogenicity	Texture
Default	0.2	2.8	3.0	0	33
Adjusted 1	0.2	3.5	3.0	0	33
Adjusted 2	0.5	3.5	2.0	0	33
Adjusted 3	0.5	2.8	2.0	0	33

3.2.1.4.3 CAD TI-RADS output

The software analysed and computed the malignancy risk category of the nodules based on 8 malignancy risk stratification systems in the software as demonstrated in Figure 3:1. To determine the diagnostic performance of the CAD software, the CAD risk stratification output for each nodule based on the 6 risk stratification systems (ACR; ATA, BTA, EU, Kwak and KSThR TIRADS) was compared to the ground truth which was the final cytological or histopathological diagnosis of each nodule. Since this study evaluated TIRADS with 5 or more malignancy risk stratification categories, the AACE/ACE/AME and the TIRADS by Seo *et al*., [267] were excluded from the analysis as they had 3 and 4 categories, respectively.

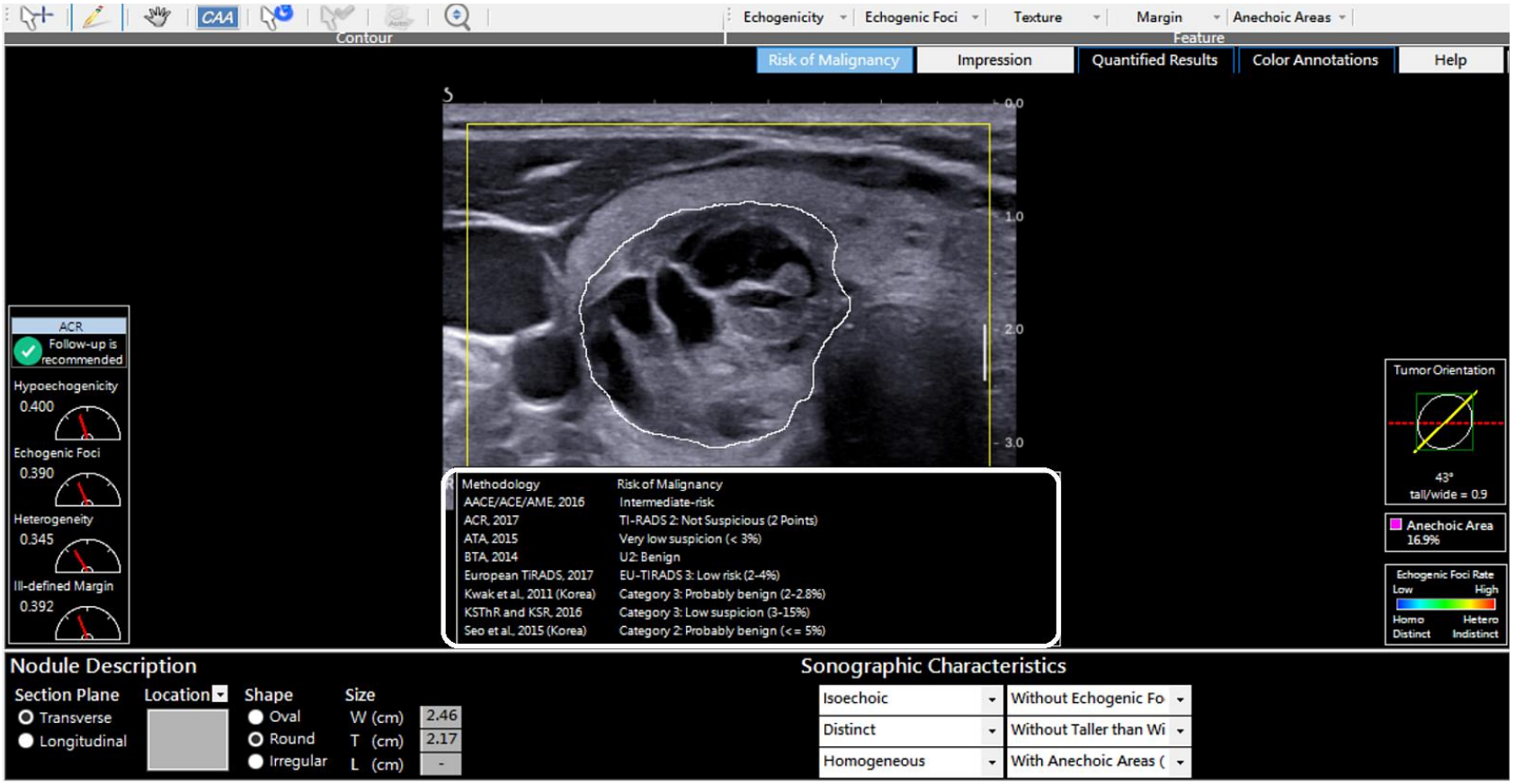


Figure 3:1: The AmCAD-UT diagnosis output of a cytologically-confirmed benign nodule is outlined in the white box as given by the category of risk of malignancy based on different TIRADS (Methodology).

3.3 Comparison of computer-assisted subjective interpretation and computer-aided diagnosis

3.3.1 Image collection procedures

For this part of the study, a consecutive case analysis approach was used for selecting the images of thyroid ultrasound imaging scans that were conducted prospectively by our research group. Standard thyroid ultrasound imaging protocols were observed to acquire 162 thyroid nodule images using a Supersonic Aixplorer ultrasound machine (SuperSonic Imagine, Aix-en-Provence, France) in conjunction with a 4 - 15MHz linear transducer. The thyroid ultrasound scans were conducted with each patient lying supine on an examination couch with the neck slightly hyperextended and the head turned away from the side of interest. Two thyroid surgeons with extensive experience conducted the ultrasound-guided fine-needle aspiration cytology (FNAC) and provided cytological and/or histopathological diagnosis of the thyroid nodules.

The inclusion criteria were like those of the software validation part of the study. However, additional criteria for this part of the study were that the nodules had to be $\geqslant$ 5mm with complete size measurements in both transverse and longitudinal planes and taller-than-wide ratio assessment and confirmatory cytological and/or histopathological results. Images of nodules < 5mm were excluded because they did not meet the size criteria for using the online calculator for assisted subjective interpretation. The reference standard was the conclusive FNAC and/or histopathology results of the nodules.

3.3.2 Analyses of the thyroid nodule images

The analyses were conducted separately with an online malignancy risk assessment system (www.gap.pe.kr/thyroidnodule.php) and AmCAD-UT version 2.2 (AmCad Biomed, Taipei, Taiwan) for the same nodules.

3.3.2.1 Computer-assisted subjective risk assessment

Two raters (R_1 and R_2) independently reviewed the same set of ultrasound images and evaluated the ultrasound features of the thyroid nodules using the stipulated scoring criteria of the online risk calculator (*Figure 3:2a*). They were both blinded to the cytology and histopathology results. The online calculator computed the malignancy risk based on a rater's subjective assessment of the composition, margins, echogenicity, shape and calcification of thyroid nodules [258]. In this study, a calculated taller-than-wide ratio of > 1 was used to determine if a nodule was taller-than-wide in addition to subjective visual assessment [268, 269]. The malignancy risk assessment was automatically computed as risk stratification category outputs for AACE, ATA, KTA/KSThR, French TIRADS and an estimated malignancy risk (EMR) score (Figure *3:2b*). In this study, the risk stratification outputs for the French TIRADS were converted to EU-TIRADS, an updated version of French TIRADS, based on the corresponding malignancy risk estimation percentages, for comparison with the AmCAD-UT EU-TIRADS outputs.

3.3.2.2 CAD assessment

The CAD analysis approach was similar to the validation aspect of the study; however, for this comparative aspect of the study, automated nodule segmentation with manual correction was used to outline the ROI. Automated nodule segmentation was selected and if the software automatically segmented the nodule boundaries satisfactorily then the computed diagnosis was accepted as valid. If the automated nodule segmentation missed the nodule, under or over-estimated the nodule boundaries, then manual segmentation was used. In this study 15 out of 162 (9.3%) images required manual correction of the automated segmentation. The malignancy risk category output for each TIRADS, taller-than-wide ratio output and sonographic characteristics outputs for each nodule on CAD (*Figure 3:3*) were compared to the

corresponding entries for each computer-assisted rater. There was no output for the EMR score in CAD hence this was not compared between the two approaches.

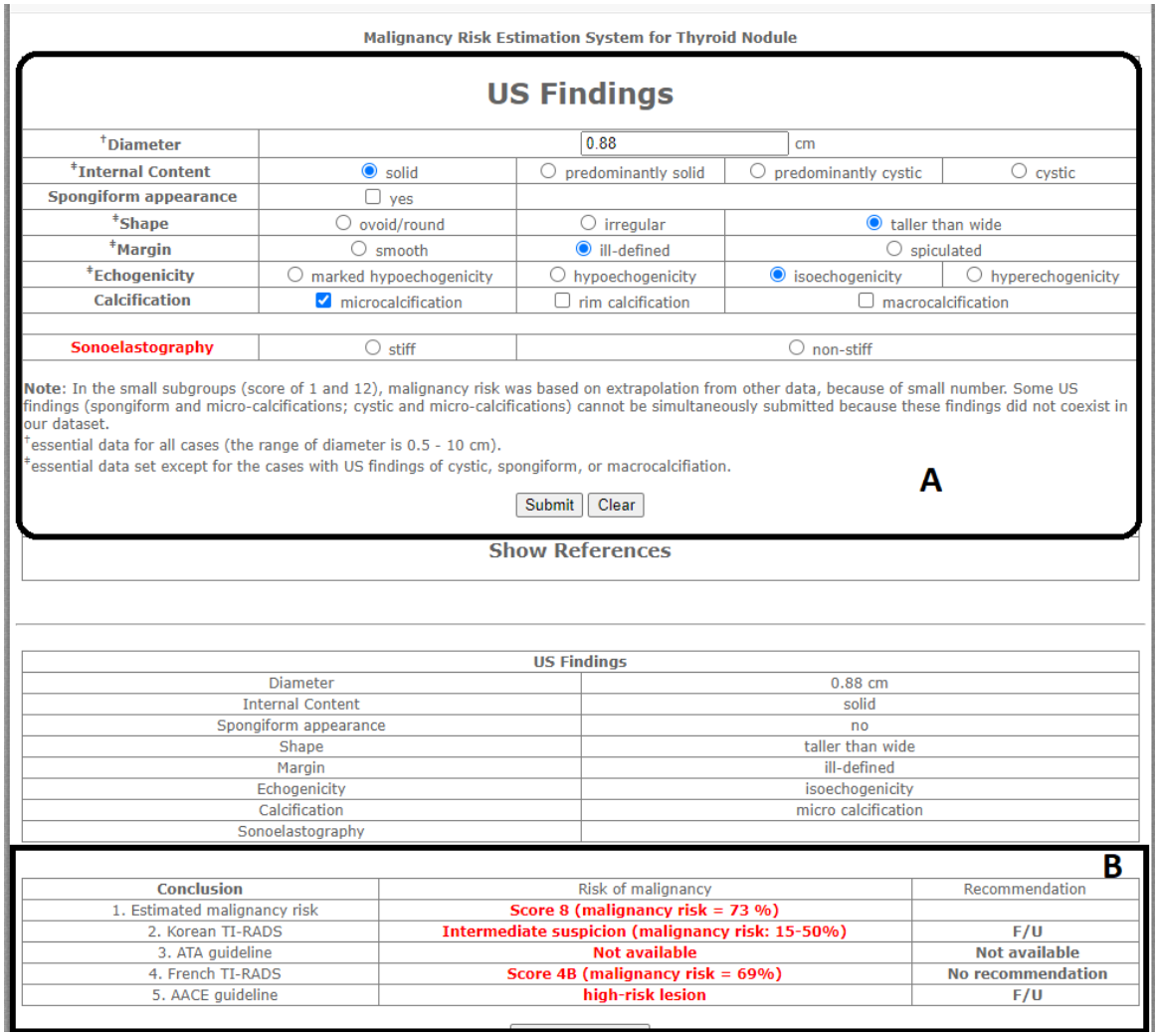


Figure 3:2: An image showing the web-based risk calculator user interface.

A: The section for sonographic characteristics inputs based on the assessment of the image by a subjective rater. B: Computed malignancy risk stratification output obtained after the submission of the sonographic feature inputs to the system. The output is based on AACE, ATA, KSThR and French TIRADS.

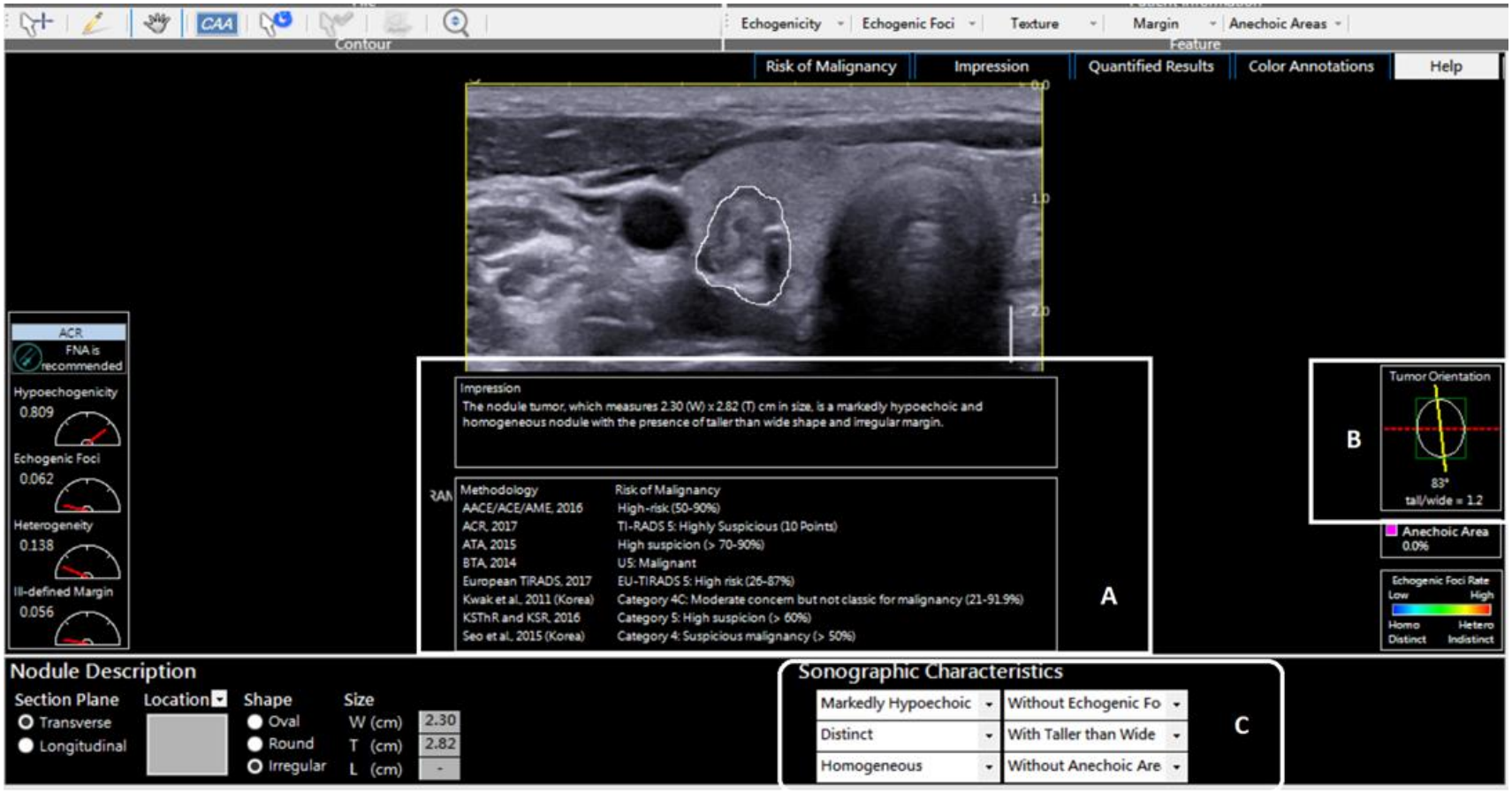


Figure 3:3: Typical CAD image analysis output that was compared to that of computer-assisted subjective rating.

A: The malignancy risk stratification of the nodule based on different classification systems. B: The software's computation of the taller-than-wide ratio. C: The software's computation of the sonographic features based on the analysed image.

3.3.3 Data analysis and statistical analysis

The statistical analysis was performed using the SPSS software package (version 26.0, SPSS Inc., Chicago, IL, USA). Categorical variables were expressed as percentages and continuous variables were expressed as mean values ± standard deviation (SD). The Chi-square test was used to compare differences in classification data while the Mann Whitney U test was used to compare continuous variables. The Shapiro Wilk test was used to check the normality of the data and based on the statistically significant output which denoted non-normality of the data, non-parametric tests were used for the statistical analyses.

In both parts of the study, the diagnostic performance measures: sensitivity (SEN), specificity (SPEC), negative predictive value (NPV), positive predictive value (PPV), diagnostic accuracy (DA) and their corresponding 95% confidence intervals (CI) were calculated with reference to final cytology and/or histopathology results. In addition to these, negative likelihood ratios (NLR), positive likelihood ratios (PLR) and diagnostic odds ratios (DOR) were calculated for the comparative analysis aspect of the study. The receiver operating characteristic (ROC) curves were generated to obtain the areas under the ROC curve (AUROC) Optimal cut-off points and equivalent diagnostic performance measures were determined from the ROCs. In the validation part of the study, precision-recall curves (PRC) were also generated to complement the ROC results for a more comprehensive diagnostic performance assessment [270]. The optimal cut-off points for all TIRADS were deemed as those that resulted in the highest Youden's index that corresponded to a compromise of both sensitivity and specificity with the least difference between the two at a higher sensitivity [271, 272].

McNemar' and Cochran Q's tests were used for the comparative analysis of SEN, SPEC, and DA whereas a two-sample proportion test was used to compare NPV and PPV. The differences in LRs and DORs were evaluated based on 95% CIs, where non-overlapping values denoted statistical significance whereas for paired comparisons of AUROCs the SPSS software

computed the differences based on the Z-test. The Goodman and Kruskal's Gamma correlation coefficient (G or γ) was used to measure the ordinal association of the different raters in the comparative analysis aspect. The gamma coefficient was interpreted as 0.01 - 0.30 negligible association, 0.31 - 0.50 low association, 0.51 - 0.7 moderate association, 0.71 - 0.9 high association and 0.91 - 1.0 very high association [273]. The inter-rater reliability testing for the different raters was estimated using Cohen's kappa statistic (κ) which were complemented by the proportion agreement. The Kappa result was interpreted as follows: 0.01 - 0.20 none to a slight agreement, 0.21 - 0.40 fair agreement, 0.41 - 0.60 moderate agreement 0.61 - 0.80 substantial agreement, and 0.81 - 1.00 almost perfect agreement [274]. All tests were two-sided and $p < 0.05$ denoted statistical significance.

3.4 Results

3.4.1 Validation of software settings

3.4.1.1 Characteristics of the thyroid nodules

A total of 205 thyroid nodules consisting of 98 (48%) malignant nodules and 107 (52%) benign nodules from 198 patients (170 females; 28 males) were included in this study. The mean age of the patients was 53.4 years ± 14.7 and the age range of the patients was 75 years (21-95). The mean short-axis diameter of malignant nodules was smaller and statistically significant from that of benign nodules (1.75 ± 0.93 cm vs 2.64 ± 1.70cm, $p < 0.05$). Thirty-five nodules (13 benign and 22 malignant) had a diameter less than 1cm. There were 77 nodules between 1 and 2cm (31 benign and 46 malignant) while 93 nodules were greater than 2cm (63 benign and 30 malignant).

3.4.1.2 Diagnostic performance of AmCAD-UT at different adjusted settings

The diagnostic performance metrics for nodules were analysed at different adjusted settings. The results are shown in Table 3.2. The optimal TIRADS cut-off point was determined to be category 4 which was the moderate suspicion with ACR, intermediate suspicion with ATA, Kwak, EU and KSThR and suspicious level with BTA TIRADS. The default setting yielded the best optimal diagnostic performance for all diagnostic measures with all TIRADS. At this setting, Kwak TIRADS had the highest AUROC (0.74) with a sensitivity of 76.6% and a specificity of 54.2%. EU TIRADS achieved the highest sensitivity and NPV and lowest specificity (SEN: 82.7%, NPV: 72.6%, SPEC: 42.1%). ATA TIRADS had the highest specificity and PPV at a sensitivity of 69.4% (SPEC: 66.4%, PPV: 65.4%). BTA TIRADS yielded the lowest sensitivity and NPV (SEN: 66.3%, NPV: 66.3%, SPEC: 60.8%) while ACR TIRADS has the lowest PPV (56.1%) with a sensitivity of 79.6% and specificity of 43%. All six TIRADS generally had similar diagnostic performance and discrimination of benign and

malignant thyroid nodules as evidenced by the ROC curve (*Figure 3:4*). The diagnostic accuracy was good based on the AUROC of 70% and above for ACR, ATA, Kwak and KSThR TIRADS. The optimal precision and recall were derived from Kwak and ATA TIRADS (*Figure 3:5*) at the chosen cut-off category. At the different cut-off points, there was high precision at low recall for all TIRADS such that even at the optimal cut-off category the PPV was substantially lower than the sensitivity.

Table 3.2: AmCAD-UT Diagnostic performance at different adjusted settings

AmCAD-UT Setting	TIRADS	SEN % (CI)	SPE (%) (CI)	[1]PPV(%) (CI)	[1]NPV% (CI)	[1]DA (%) (CI)	AUROC (CI)
Default	ACR-4	79.6 (70.3; 87.1)	43.0 (33.5 ; 52.9)	56.1 (51.3 ; 60.8)	69.7 (59.5 ; 78.3)	60.5 (53.4 ; 67.2)	.72 (.65 ; .79)
	ATA-4	69.4 (59.3 ; 78.3)	66.4 (56.6 ; 75.2)	65.4 (58.4 ; 71.8)	70.3 (63.1 ; 76.7)	67.8 60.9 ; 74.1)	.73 (.66 ; .80)
	BTA-4	66.3 (56.1; 75.6)	60.8 (50.8 ; 70.1)	60.8 (54.0 ; 67.1)	66.3 (58.9 ; 73.0)	63.4 (56.4 ; 70.0)	.66 (.59 ; .74)
	EU-4	82.7 (73.7 ; 89.6)	42.1 (32.6 ; 52.0)	56.6(52.1 ; 61.1)	72.6 (62.0 ; 81.2)	61.5 (54.4 to 68.2)	.65 (.58 ; .71)
	Kwak-4b	76.5 (66.9 ; 84.5)	54.2 (44.3; 63.9)	60.5 (54.8 ; 65.9)	71.6 (62.9 ; 79.0)	64.5 (57.9 ; 71.4)	.74 (.67 ; .80)
	KSThR-4	77.6 (68.0 ; 85.4)	50.5 (40.6 ; 60.3)	58.9 (53.5 ; 64.1)	71.1 (61.9 ; 78.8)	63.4 (56.4 ; 70.0)	.70 (.63 ; .76)
Adjusted 1	ACR-4	74.5 (64.7 ; 82.8)	48.6 (38.8 ; 58.5)	57.0 (51.6 ; 62.3)	67.5 (58.5 ; 75.5)	61.0 (53.9 ; 67.7)	.67 (.60 ; .74)
	ATA-4	70.4 (60.3 ; 79.2)	67.3 (57.6 ; 76.1)	66.4 (59.3 ; 72.7)	71.3 (64.0 ; 77.6)	68.8 (62.0 ; 75.1)	.71 (.64 ; .78)
	BTA-4	69.4 (59.3 ; 78.3)	58.9 (49.9 ; 68.3)	60.7 (54.3; 66.8)	67.7 (60.0 ; 74.6)	63.9 (56.9 ; 70.5)	.67 (.59 ; .74)
	EU-4	81.6 (72.5 ; 88.7)	46.7 (37.0 ; 56.6)	58.4 (53.5 ; 63.2)	73.5 (63.6 ; 81.5)	63.4 (56.4 ; 70.0)	.66 (.59 ; .73)
	Kwak-4b	71.4 (61.4 ; 80.1)	57.0 (47.1 ; 66.5)	60.3 (54.2 ; 66.2)	68.5 (60.5 ; 75.6)	63.9 (56.9 ; 70.5)	.68 (.61 ; .75)
	KSThR-4	71.4 (61.4 ; 80.1)	58.9 (49.0 ; 68.3)	61.4 (55.1 ; 67.3)	69.2 (61.3 ; 76.2)	64.9 (57.9 ; 71.4)	.70 (.63 ; .76)
Adjusted2	ACR-4	61.2 (50.9 ; 70.9)	57.9 (48.0 ; 67.4)	57.1 (50.4 ; 63.7)	62.0 (54.8 ; 68.7)	59.5 (52.5 ; 66.3)	.62 (.55 ; .70)
	ATA-4	44.9 (34.8 ; 55.3)	78.5 (69.5 ; 85.9)	65.7 (55.6 ; 74.5)	60.9 (55.9 ; 65.6)	62.4 (55.4 ; 69.1)	.59 (.51 ; .67)
	BTA-4	56.1 (45.7 ; 66.1)	64.5 (54.7 ; 73.5)	59.1 (51.5 ; 66.4)	61.6 (55.2 ; 67.6)	60.5 (53.4 ; 67.2)	.60 (.53 ; .68)
	EU-4	71.4 (61.4 ; 80.1)	55.1 (45.2 ; 64.8)	59.3 (53.3 ; 65.1)	67.8 (59.6 ; 75.1)	62.9 (55.9 ; 69.6)	.65 (.57 ; .72)
	Kwak-4b	50.0 (39.7 ; 60.3)	68.2 (58.5 ; 76.9)	59.0 (50.6 ; 67.0)	59.8 (54.1 ; 65.4)	59.5 (52.5 ; 66.3)	.61 (.54 ; .68)
	KSThR-4	59.2 (48.8 ; 69.0)	62.6 (52.7 ; 71.8)	59.2 (51.9 ;66.1)	62.6 (55.9 ; 68.9)	61.0 (53.9 ; 67.7)	.62 (.55 ; .69)
Adjusted3	ACR-4	78.5 (69.1 ; 86.2)	41.1 (31.7 ; 51.1)	55.0 (50.3 ; 59.6)	67.7 (57.4 ; 76.5)	59.0 (52.0 ; 65.8)	.67 (.60 ; .74)
	ATA-4	44.9 (34.8 ; 55.3)	78.5 (69.5 ; 85.9)	65.7 (55.6 ; 74.5)	60.9 (55.9 ; 65.6)	62.4 (55.4 ; 69.1)	.58 (.50 ; .66)
	BTA-4	56.1 (45.7 ; 66.1)	64.5 (54.7 ; 73.5)	59.1 (51.5 ; 66.4)	61.6 (55.2 ; 67.6)	60.5 (53.4 ; 67.2)	.62 (.55 ; .70)
	EU-4	82.7 (73.7 ; 89.6)	42.1 (32.6 ; 52.0)	56.6 (52.1 ; 61.1)	72.6 (62.0 ; 81.2)	61.5 (54.4 ; 68.2)	.63 (.56 ; .69)
	Kwak-4b	63.3 (52.9 ; 72.8)	57.0 (47.1 ; 66.5)	57.4 (50.8 ; 63.7)	62.9 (55.5 ; 69.7)	60.0 (53.0 ; 66.8)	.64 (.57 ; .71)
	KSThR-4	77.6 (68.0 ; 85.4)	43.9 (34.3 ; 53.9)	55.9 (50.9 ; 60.7)	68.1 (58.3 ; 76.6)	60.0 (53.0 ; 66.8)	.63 (.57 ; .70)

[1]PPV, NPV and DA values were not calculated based on prevalence

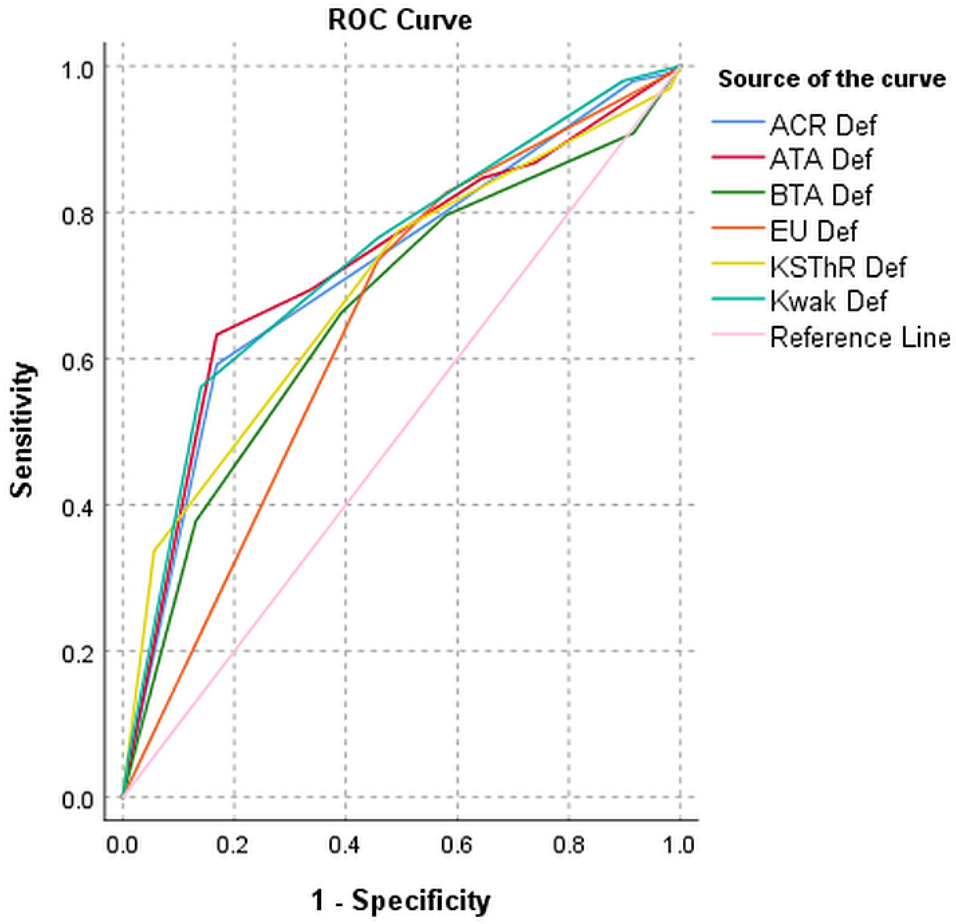


Figure 3:4: ROC curve demonstrating the diagnostic performance of the 6 TIRADS at the default setting

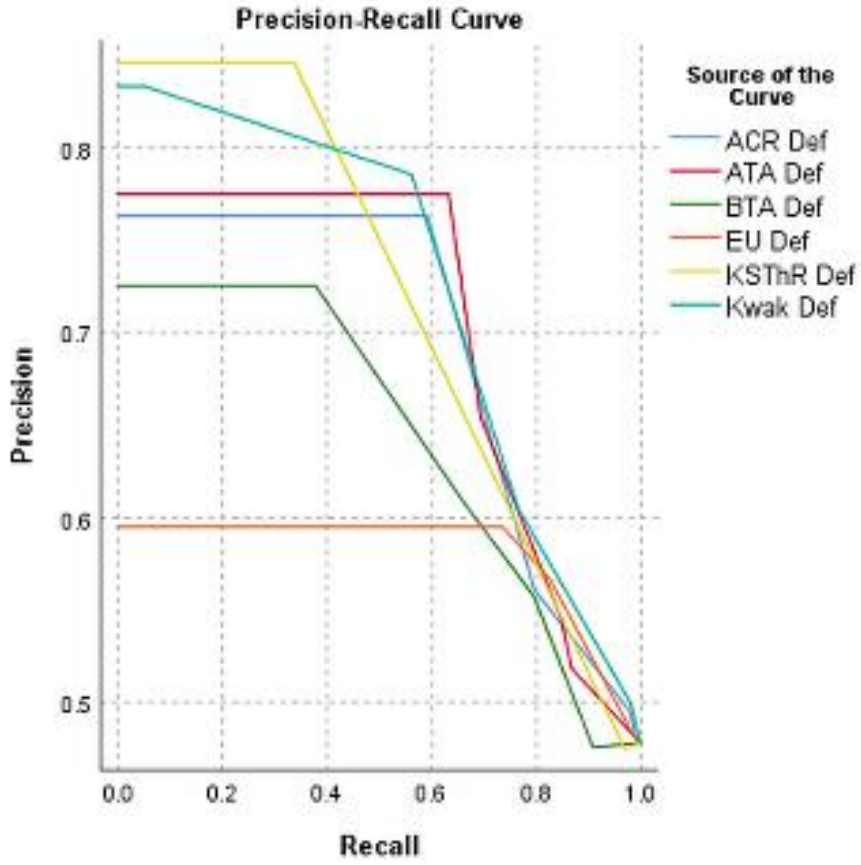


Figure 3:5: PRC curve demonstrating the diagnostic performance of the 6 TIRADS at the default setting

For all TIRADS, the default setting and Adjusted 1 setting had comparable diagnostic performances whereas Adjusted 2 and Adjusted 3 had lower diagnostic performances (*Table 3.2*). Exclusively lowering the sensitivity to hyperechoic foci to 3.5 (Adjusted 1), resulted in a slight incremental change in sensitivity and specificity and a good AUROC (0.71) with ATA TIRADS, however, this was not statistically significant. The other TIRADS had a slightly lower sensitivity and AUROC with a slight increase in specificity except for KSThR which maintained an AUROC of 0.7. Conversely, the sensitivity of BTA was increased while the specificity was reduced (SEN: 69.4%, SPEC: 58.9%). The AUROC at the adjusted settings was generally lower than at the default setting for all TIRADS.

The differences amongst the 6 TIRADS at different settings are illustrated in Table 3.3. At the default setting, there were significant differences in sensitivity between EU and 2 TIRADS (BTA and ATA, $p < 0.05$). For specificity, the most significant differences were between ACR and KSThR; ACR and BTA and ATA and EU ($p < 0.001$). Statistically significant differences in AUROC were mostly between BTA and Kwak, and EU and Kwak ($p < 0.001$). There were no significant differences between the AUROC of different TIRADS at all adjusted settings ($p > 0.05$). The most significant differences for both sensitivity and specificity were between ATA and EU TIRADS at all adjusted settings ($p < 0.05$). The post-adjustment diagnostic performance results of EU TIRADS did not differ from those of the default setting ($p > 0.05$), except at Adjusted 2. ACR and EU TIRADS sensitivity and specificity had no statistically significant differences in all settings ($p > 0.05$).

Table 3.3:Pairwise comparison of TIRADS diagnostic performance in CAD-differentiation of malignant and benign thyroid nodules at different settings

TIRADS Pairs	**Performance measures (p-values)**											
	Default			**Adjusted 1**			**Adjusted 2**			**Adjusted 3**		
	SEN*	**SPEC***	**AUROC**	**SEN***	**SPEC***	**AUROC**	**SEN***	**SPEC***	**AUROC**	**SEN***	**SPEC***	**AUROC**
ATA - ACR	.04	<.0001	.826	.999	<.0001	.262	.003	<.0001	.546	<.0001	<.0001	.079
ATA - BTA	.999	.999	.158	.999	.605	.951	.151	.004	.393	.305	.027	.080
ATA - EU	.001	<.0001	.051	.001	<.0001	.873	< .0001	<.0001	.224	< .0001	<.0001	.057
ATA - Kwak	.539	.054	.795	.999	.183	.391	.999	.120	.497	.002	< .0001	.120
ATA - KSThR	.248	.002	.409	.999	.605	.094	.016	.001	.992	<.0001	< .0001	.073
ACR - BTA	.001	<.0001	.019	.999	.183	.338	.999	.999	.814	<.0001	<.0001	.427
ACR - EU	.999	.999	.001	.530	.999	.268	.289	.999	.310	.999	.999	.377
ACR - Kwak	.999	.108	.289	.999	.605	.490	.151	.120	.686	.023	.006	.204
ACR - KSThR	.013	<.0001	.248	.999	.183	.732	.999	.999	.554	.999	.999	.284
BTA - EU	.442	<.0001	.530	.005*	.046	.860	.007	.238	.091	<.0001	<.0001	.865
BTA - Kwak	.07*	.018	<.0001	.999*	.999	.486	.999	.999	.818	.999	.999	.522
BTA - KSThR	.206	.206	.168	.999	.999	.115	.999	.999	.343	<.0001	<.0001	.712
EU - Kwak	.999	.054	<.0001	.040	.183	.472	< .0001	.011	.159	.001	.013	.694
EU - KSThR	.999	.657	.022	.040	.046	.140	.075	.806	.169	.999	.999	.720
Kwak - KSThR	.999	.999	.030	.999	.999	.411	.528	.999	.604	.047	.053	.816

3.4.1.3 CAD sonographic feature outputs of false negative and false nodules

The different AmCAD-UT settings consistently misdiagnosed 48 nodules (32 false positives and 16 false negatives) across all TIRADS. Based on the software's sonographic interpretation of analysed nodules, 78% (25/32) of the false-positive nodules had a homogenous echotexture, 75% (24/32) were predominantly solid and 53% (17/32) had irregular margins. Moreover, 59% (19/32) of false-positive nodules had no calcifications, whereas 6 out of the remaining 13 false-positive nodules (46%) had detected echogenic foci attributed to mixed calcifications. The distribution of the echogenicity feature varied, with a combined 44% total for hypoechoic and markedly hypoechoic (7/32 and 7/32 respectively). All 16 false-negative nodules had regular margins, 75% (12/16) were heterogeneous, 56% (9/16) were hyperechoic and 38% (6/16) had microcalcifications. Examples of nodules that the CAD misdiagnosed based on its detection of sonographic features are demonstrated in Figure 3:6 and Figure 3:7.

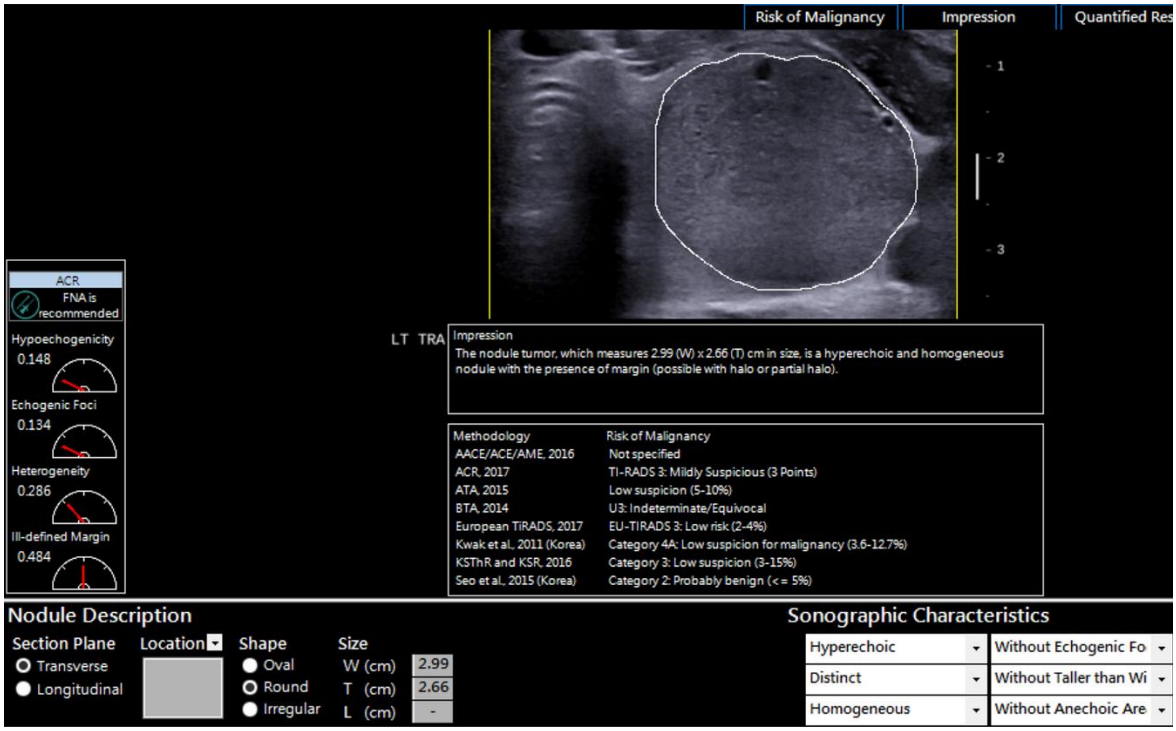


Figure 3:6: An example of a thyroid nodule misdiagnosed as false negative for malignancy. The misdiagnosis was based on the software's impression of the hyperechoic and homogenous sonographic features. The histopathological diagnosis of the nodule was papillary thyroid cancer.

Figure 3:7: An example of a solid markedly hypoechoic thyroid nodule that was diagnosed as false positive.

The nodule was also taller than wide with echogenic foci. The impression of these suspicious sonographic features led to the misdiagnosis by the different TIRADS within the AmCAD-UT software. The nodule was diagnosed as benign with cytology.

3.4.2 Comparative analysis of CAD and computer-assisted subjective interpretation

3.4.2.1 Demographics and thyroid nodule characteristics

This part of the study included 162 thyroid nodules from 157 patients of which 133 (84.7%) were females and 24 (15.3%) were males. The mean age of the patients was 53 years ± 13 (range 21- 95 years). There were 100 (61.7%) benign thyroid nodules and 62 (38.3%) malignant nodules. A total of 41 (41%) benign nodules had benign cytology results whereas 27 (27%) had benign histopathology findings without defined pathology while the rest had pathological findings of goitre (n=14), adenomas (n=6), hyperplasia (n=3), colloid nodule (n=1), lymphocytic thyroiditis (n=2), Hashimoto's thyroiditis (n=2) and Graves' disease (n=4). There 54 (87%) malignant thyroid nodules were papillary thyroid carcinoma while the rest were follicular thyroid carcinoma (n=5), Hurthle cell carcinoma (n=1) and non-invasive follicular thyroid neoplasm with papillary-like nuclear features-NIFTP (n=2). The mean diameter of all thyroid nodules was 1.57 ± 0.84 cm (range 0.5-3.9 cm). The mean diameter did not differ significantly between benign and malignant thyroid nodules (1.63 ± 0.85 cm and 1.46 ± 0.83 cm respectively, $p = 0.15$).

3.4.2.2 Nodule sonographic feature classifications by human subjective assessment and CAD

The two subjective raters and the CAD system coded the sonographic features of the thyroid nodules based on echogenicity, composition, calcifications, margins, and shape. Table 3.4 shows the results of the different categorisations. There were significant differences in all sonographic feature classifications between benign and malignant nodules for both subjective raters ($p < 0.05$) whereas significant differences between benign and malignant nodules were only observed for calcifications ($p = 0.001$) for the CAD system. Amongst both human raters and CAD, majority of the malignant nodules were classified as hypoechoic compared to benign nodules (R_1 = 66.1% vs 48%, R_2 = 72.6% vs 49% and CAD = 48.4% vs 34%). All raters

classified majority of the malignant nodules as either predominantly solid or solid (R_1 = 16.1% and 79%, R_2 = 54.8% and 41.9% and CAD = 38.7% and 61.3%, respectively). Microcalcifications were predominant in malignant nodules when classified by R_1 and CAD (46.8% and 51.6%, respectively), but R_2 classified malignant nodules as mostly without calcifications or with microcalcifications (both 45.2%).

3.4.2.3 Classification correlation comparisons between subjective ratings and CAD

There was a very high association in the rating of echogenicity of malignant nodules between both human raters and a high association between each human rater and CAD (R_1 vs R_2, G = 0.91; R_1 vs CAD, G = 0.74; and R_2 vs CAD, G = 0.73) as shown in Table 3.5. The subjective raters stratified calcifications and composition in all nodules and separate groups of malignant nodules and benign nodules with high association (G > 0.7). There was negligible to low association between each of the human raters and the CAD for classifying nodule margins in malignant, benign and all total nodules (G < 0.5). There was generally high rank correlation association in categorising the shape of benign nodules between the human raters and between each human rater and CAD (R_1 and R_2, G = 0.85; R_1 vs CAD, G = 0.81 and R_2 vs CAD, G = 0.86; respectively).

Table 3.4: Classifications of sonographic features for differentiating benign and malignant nodules based on the ratings of the 2 raters and CAD

Sonographic		**R_1**			**R_2**			**CAD**		**P- values**		
feature	M = 62	B = 100	T = 162	M = 62	B = 100	T= 162	M = 62	B = 100	T = 162	R_1	R_2	CAD
Echogenicity												
Isoechoic	9 (14.5%)	35 (35%)	44 (27.2%)	12 (19.4%)	45 (45%)	57 (35.2%)	11 (17.7%)	27 (27%)	38 (23.5%)	0.006	<0.001	0.13
Hyperechoic	5 (8.1%)	13 (13%)	18 (11.1%)	0 (0%)	5 (5%)	5 (3.1%)	13 (21%)	31 (31%)	44 (27.2%)			
Hypoechoic	41 (66.1%)	48 (48%)	89 (54.9%)	45 (72.6%)	49 (49%)	94 (58%)	30 (48.4%)	34 (34%)	64 (39.5%)			
M-hypoechoic	7 (11.3%)	4 (4%)	11 (6.8%)	5 (8.1%)	1 (1%)	6 (3.7%)	8 (12.9%)	8 (8%)	16 (9.9%)			
Calcifications												
none	18 (29%)	61 (61%	79 (48.8%)	28 (45.2%)	78 (78%)	106 (65.4%)	27 (43.5%)	71 (71%)	98 (60.5%)	<0.001	<0.001	0.001
macrocalcification	2 (3.2%)	10 (10%)	12 (7.4%)	2 (3.2%)	2 (2%)	4 (2.5%)	0 (0%)	0 (0%)	0 (0%)			
microcalcification	29 (46.8%)	25 (25%)	54 (33.3%)	28 (45.2%)	19 (19%)	47 (29%)	32 (51.6%)	23 (23%)	55 (34%)			
rim calcification	1 (1.6%)	1 (1%)	2 (1.2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)			
mixed calcification	12 (19.4%)	3 (3%)	15 (9.3%)	4 (6.5%)	1 (1%)	5 (3.1%)	3 (4.8%)	6 (6%)	9 (5.6%)			
Margins												
well-defined	26 (41.9%)	71 (71%)	97 (59.9%)	31 (50%)	70 (70%)	101 (62.3%)	42 (67.7%)	80 (80%)	122 (75.3%)	0.001	0.001	0.093
irregular	32 (51.6%)	26 (26%)	58 (35.8%)	18 (29%)	26 (26%)	44 (27.2%)	20 (32.3%)	20 (20%)	40 (24.7%)			
microlobulated	0 (0%)	0 (0%)	0 (0%)	1 (1.6%)	2 (2%)	3 (1.9%)	0 (0%)	0 (0%)	0 (0%)			
spiculated	4 (6.5%)	3 (3%)	7 (4.3%)	12 (19.4%)	2 (2%)	14 (8.6%)	0 (0%)	0 (0%)	0 (0%)			

Sonographic	R_1			R_2			CAD			P- values		
feature	M = 62	B = 100	T = 162	M = 62	B = 100	T= 162	M = 62	B = 100	T = 162	R_1	R_2	CAD
Composition												
spongiform	3 (4.8%)	15 (15%)	18 (11.1%)	1 (1.6%)	13 (13%)	14 (8.6%)	0 (0%)	1 (1%)	1 (0.6%)	0.005	0.014	0.90
predominantly cystic	0 (0%)	7 (7%)	7 (4.3%)	1 (1.6%)	7 (7%)	8 (4.9%)	0 (0%)	1 (1%)	1 (0.6%)			
predominantly solid	10 (16.1%)	23 (23%)	33 (20.4%)	34 (54.8%)	51 (51%)	85 (52.5%)	24 (38.7%)	35 (35%)	59 (36.4%)			
solid	49 (79%)	55 (55%)	104 (64.2%)	26 (41.9%)	29 (29%)	55 (34%)	38 (61.3%)	63 (63%)	101 (62.3%)			
Shape												
Round/Ovoid	25 (40.3%)	74 (74%)	99 (61.1%)	28 (45.2%)	75 (75%)	103 (63.6%)	49 (79%)	90 (90%)	139 (85.8%)	<0.001	<0.001	0.057
Taller than wide	28 (45.2%)	20 (20%)	48 (29.6%)	27 (43.5%)	21 (21%)	48 (29.6%)	13 (21%)	9 (9%)	22 (13.6%)			
Irregular	9 (14.5%)	6 (6%)	15 (9.3%)	7 (11.3%)	4 (4%)	11 (6.8%)	0 (0%)	1 (1%)	1 (0.6%)			

M = malignant nodules, B = benign nodules, T = total nodules, M-hypoechoic = markedly hypoechoic R1 = Rater1, R2 = Rater 2, CAD = AMCAD-UT.

Table 3.5: Rank correlation (γ) of rating of sonographic features by the 2 raters and CAD.

Sonographic feature	**R_1 vs CAD**			**R_2 vs CAD**			**R_1 vs R_2**		
	M	**B**	**T**	**M**	**B**	**T**	**M**	**B**	**T**
Echogenicity	0.74	0.55	0.64	0.73	0.63	0.67	0.91	0.64	0.78
Calcifications	0.30	0.45	0.49	0.33	0.26	0.41	0.84	0.78	0.85
Margins	0.39	0.14	0.33	0.37	0.22	0.35	0.03	0.63	0.45
Composition	-0.25	0.46	0.28	0.15	0.60	0.46	0.81	0.79	0.83
Shape	0.65	0.81	0.76	0.40	0.86	0.71	0.41	0.85	0.72

M = malignant nodules, B = benign nodules, T = all total nodules, R_1 = Rater1, R_2 = Rater 2, CAD = AMCAD-UT

3.4.2.4 Rater agreement based on TIRADS

The rater agreement was determined based on the malignancy cut-off points of the different TIRADS. The results are shown in Table 3.6. Agreement between the human raters was generally moderate to substantial for malignant, benign and all total nodules ($0.41 \leq \kappa < 0.81$) and the highest agreement was achieved with ATA TIRADS ($\kappa = 0.77$). The rater agreement between each of the human raters and the CAD was highest based on ATA TIRADS between R_1 and CAD for all nodules ($\kappa = 0.75$), and lowest based on AACE for malignant nodules between R_1 and CAD and R_2 and CAD ($\kappa = 0.12$ and $\kappa = 0.14$, respectively). There was a fair rate of agreement for classifying benign nodules with AACE ($\kappa = 0.32$) between R_1 and CAD, and with ATA, EU and KSThR ($\kappa = 0.40$, 0.24 and 0.23, respectively) between R_2 and CAD. Proportions of agreement between the different paired raters and CAD amongst all TIRADS were generally high in contrast to the moderate kappa values, although the agreement between R_2 vs CAD was low to moderate for benign nodules with all TIRADS -AACE = 50.8%, ATA = 73.7%, EU = 63%, and KSThR = 61%. (**Appendix 1**).

3.4.2.5 Diagnostic performance assessment of CAD and computer-assisted raters for matched TIRADS

The diagnostic performance outcomes for the 2 computer-assisted subjective raters and CAD were assessed for different TIRADS as outlined in Table 3.7. The best diagnostic performance for the different TIRADS was achieved at high risk (category 3) for AACE, high suspicion (category 5) for ATA and EU, and intermediate suspicion (category 4) for KSThR. EU and KSThR TIRADS were able to specify all nodules, regardless of the rater whereas AACE rating with CAD failed to specify some nodules (39 benign, 9 malignant), and ATA failed to specify some nodules regardless of the rater (CAD – 30 benign, 19 malignant; R_1 – 16 benign, 10 malignant; and R_2 – 15 benign, malignant). Overall, the common nodules across all raters that

could be specified by AACE, and ATA were 114 (61 benign, 53 malignant) and 96 (57 benign, 39 malignant), respectively.

Based on the different TIRADS, CAD yielded the highest sensitivity but lowest specificity and AUROC amongst the two human raters and CAD with AACE (92.5%; 26.2% and 0.59, respectively) which were all statistically significant different from R_2 (79.3%, 62.3% and 0.72, $p < 0.05$). For stratifying all 162 nodules, R_1 had overall higher diagnostic performance than CAD for all metrics for EU and KSThR. Although such differences in diagnostic performance were not statistically significant for EU, there was a statistically significant difference in AUROC for KSThR (CAD - 0.67 (95% CI: 0.59; 0.75) vs R_1 - 0.76 (95% CI: 0.68; 0.83), $p = 0.02$) (**Appendix 2**). R_2 had comparable sensitivity but higher specificity than CAD for KSThR (75.8% vs 83.9%, and 61% vs 46%, $p = 0.02$, respectively). Between the 2 computer-assisted subjective raters there were statistically significant differences in sensitivity, but comparable specificity and AUROCs for both EU (85.5% vs 71%, $p = 0.04$; 62% vs 64%, $p > 0.05$; and 0.74 vs 0.69, $p > 0.05$ respectively) and KSThR (90.3% vs 75.8%, $p = 0.01$; 51% vs 61%, $p > 0.05$; and 0.76 vs 0.74, $p > 0.05$ respectively). Overall, CAD generally had lower PLRs, although these were comparable to those of the computer-assisted raters, while the lowest NLR was achieved with a computer-assisted rating (KSThR- R_1 – 0.19). The highest specificity and PLR across all raters/CAD was achieved with ATA with comparable sensitivity, specificity, AUROC amongst all raters/CAD. At the best performance, the computer-assisted subjective rating approach had higher DOR > 9 and higher AUROC > 0.7 than the CAD-based approach on all the TIRADS. Across all TIRADS, all raters/CAD yielded high sensitivity and high NPVs, but low to moderate SPEC, PPVs and DAs (**Appendix 3)**.

Table 3.6: Rater agreement (κ) for the TIRADS classifications

Raters	Nodules	TIRADS			
		AACE	**ATA**	**EU**	**KSThR**
R_1 vs CAD	M	0.12	0.69	0.23	0.43
	B	0.32	0.68	0.40	0.46
	ALL	0.35	0.75	0.46	0.53
R_2 vs CAD	M	0.18	0.57	0.45	0.45
	B	0.12	0.40	0.24	0.23
	ALL	0.21	0.56	0.38	0.37
R_1 vs R_2	M	0.52	0.71	0.40	0.50
	B	0.60	0.71	0.57	0.43
	ALL	0.65	0.77	0.59	0.54

ATA= American Thyroid Association, AACE = American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi, EU- European Union, KSThR - Korean Society of Thyroid Radiology

Table 3.7: Diagnostic performance metrics of the 2 raters and CAD based on different TIRADS

RATER BY TIRADS	**TOTAL NODULES**	**SEN % (CI)**	**SPE (%) (CI)**	**PLR (CI)**	**NLR (CI)**	**DOR (CI)**	**AUROC (CI)**
EU-CAD	162	79.0 (66.8;88.3)	55.0 (44.7;65.0)	1.76 (1.37;2.26)	0.38 (0.23;0.64)	4.61 (2.23;9.53)	0.67 (0.59;0.74)
EU-R_1	162	85.5 (74.2;93.1)	62.0 (51.8;71.5)	2.25 (1.72;2.95)	0.23 (0.12;0.42)	9.61 (4.26;21.68)	0.74 (0.68;0.81)
EU-R_2	162	71.0 (58.1;81.8)	64.0 (53.8;73.4)	1.97 (1.45;2.68)	0.45 (0.30;0.69)	4.35 (2.16;8.37)	0.69 (0.61;0.76)
KSThR- CAD	162	83.9 (72.3;92.0)	46.0 (36.0;56.3)	1.55 (1.26;1.92)	0.35 (0.19;0.64)	4.43 (2.03;9.69)	0.67 (0.59;0.75)
KSThR-R_1	162	90.3 (80.1;96.4)	51.0 (40.8;61.4)	1.84 (1.49;2.29)	0.19 (0.09;0.42)	9.71 (3.84;24.59)	0.76 (0.68;0.84)
KSThR-R_2	162	75.8 (63.3;85.8)	61.0 (50.7;70.6)	1.94 (1.47;2.58)	0.40 (0.25;0.63)	4.90 (2.42;9.93)	0.74 (0.66;0.81)
AACE-CAD	114	92.5 (81.8;97.9)	26.2 (15.8;39.1)	1.25 (1.06;1.48)	0.29 (0.10;0.81)	4.36 (1.35;14.01)	0.59 (0.49;0.70)
AACE-R1	114	88.7 (77.0;95.7)	54.1 (40.9;66.9)	1.93 (1.45;2.58)	0.21 (0.10;0.46)	9.23 (3.44;24.79)	0.72 (0.63;0.81)
AACE-R_2	114	79.3 (65.9;89.2)	62.3 (49.0;74.4)	2.10 (1.48;2.98)	0.33 (0.19;0.58)	6.31 (2.72;14.64)	0.72 (0.63;0.82)
ATA-CAD	96	79.5 (63.5;90.7)	66.7 (52.9;78.6)	2.38 (1.60;3.56)	0.31 (0.16;0.59)	7.75 (2.99;20.09)	0.74 (0.63;0.84)
ATA-R_1	96	79.5 (63.5;90.7)	70.2 (56.6;81.6)	2.67 (1.70;4.19)	0.29 (0.15;0.55)	9.12 (3.48;23.87)	0.74 (0.64;0.85)
ATA-R_2	96	74.4 (57.9;87.0)	68.4 (54.8;80.1)	2.35 (1.54;3.60)	0.37 (0.21;0.66)	6.28 (2.53;15.61)	0.73 (0.63;0.83)

SEN = sensitivity, SPEC = specificity, PLR = positive likelihood ratio, NLR = negative likelihood ratio, DOR = diagnostic odds ratio, AUROC = area under receiver operator characteristic curve, CI =95% confidence interval

3.5 Discussion

3.5.1 Validation of software settings

CAD approaches have been suggested as more objective methods of ultrasound feature assessment that can perform comparatively accurate to expert subjective assessment of thyroid nodule ultrasound features. The diagnostic performance of adjusted settings of AmCAD-UT in comparison with the default setting for thyroid nodule differentiation based on six TIRADS was evaluated in this study. The results showed that the diagnostic performance of AmCAD-UT was higher at the default setting than at different adjusted settings. Although the diagnostic performance metrics were comparable amongst the six TIRADS, EU TIRADS had the highest sensitivity (82.7%); ATA the highest specificity (66.4%) while ATA, ACR, KSThR and Kwak all had good AUROC with Kwak being the highest (0.74). The results of this comparative study among six TIRADS at default and adjusted settings in AmCAD-UT are reported for the first time.

The high sensitivity and NPV of EU TIRADS; a high AUROC with Kwak TIRADS and comparable diagnostic efficiency among ACR, ATA, Kwak and KSThR TIRADS with high sensitivities have been established in human subjective assessment studies[133, 140, 275, 276]. In CAD assessment, Reverter et al. [242] evaluated the performance of AmCAD-UT at default settings using ATA, EU and AACE/ACE/AME TIRADS. ATA TIRADS had the highest diagnostic performance overall, and a sensitivity comparable to that of an experienced subjective assessor using ATA TIRADS in that study. Contrarily, this present study had a lower sensitivity for ATA than that study (SEN: 69.4% vs 87%) but with a comparable specificity in (SPEC: 66.4% vs 68.8%) as well as comparable EU TIRADS sensitivity and specificity (SEN: 82.7% vs 85.2%; SPEC: 42.1% vs 50.2%). The differences in the results of ATA sensitivity between our study and that study could be attributed to the differences in the samples, study design and TIRADS cut-off category criteria.

The evaluation of adjusted settings in our study showed that the Adjusted 1 setting had comparable results with the default setting. However, there was a marginal improvement in specificity and sensitivity for ATA TIRADS, while for most TIRADS there was a minimal increase in specificity. The differences in diagnostic performance upon the adjustment of settings can be accounted for by the diverse malignancy risk stratification criteria for the different TIRADS based on pattern-based approaches (ATA, EU, BTA and KSThR) or score-based approaches (ACR and Kwak) [132, 133, 277]. Inconsistencies among the different TIRADS are mainly in the categorisation of echogenic foci and echogenicity [278]. ATA tends to fail to classify some nodules with mixed calcifications, whereas Kwak interprets them as having microcalcifications thereby resulting in a higher fitted malignancy probability for calcifications [279, 280]. The sole reduction of the sensitivity detection of hyperechoic foci likely interfered with the detection of subtle calcifications for CAD malignancy risk stratification thereby slightly improving the specificity at the expense of lower overall sensitivity. Owing to this, the adjustment of AmCAD-UT sensitivity detection may be advisable for analysing individual problematic suspicious sonographic features that influence the malignancy risk estimation based on the selected TIRADS. For example, hyperechoic foci and anechoic areas settings may be adjusted separately in a hypoechoic nodule with mixed echogenic foci without other suggestive features or clinical correlation history. In such a case, using the Adjusted 1 setting would likely improve diagnostic accuracy. Focusing on calcifications and hypoechogenicity features in isolation may ascertain the extent to which each feature affects the CAD diagnosis output based on the different TIRADS classification divergences.

The sensitivity at the default and Adjusted 1 setting using Kwak, KSThR, and ATA TIRADS was comparable to a previous CAD study's findings for less experienced ultrasound users, while EU TIRADS had higher sensitivity (82.7%) than that same study which yielded an average sensitivity of about 72% [281]. For the ACR TIRADS category, our study had similar

sensitivity to that of a different CAD development study for sole CAD and a junior radiologist using CAD (79.6% vs 80.6% vs 78.1%, respectively). However, that study had a higher diagnostic performance for all other measures [236]. Similarly, at the default setting, the AUROC of above 0.70 with Kwak, KSThR, ACR and ATA TIRADS in our study was comparable to that of a recent multicentre and multi-reader AmCAD-UT study which averaged 0.79 regardless of user experience [252]. However, the multi-reader study outcomes were not stated as specific to any TIRADS.

In this current study, most of the misdiagnosed nodules had some typical sonographic features of suspicion for malignancy or benignity. The CAD's interpretations of solid, homogenous nodules with irregular margins and/or echogenic foci features (such as the presence of colloid) were likely the key contributors to the false-positive diagnosis of some benign nodules. This may be likely due to the high thyroid malignancy risk estimation when multiple suspicious features are present as previously determined in several non-CAD studies [4, 15, 282]. Echogenic foci with a comet-tail artefact, particularly in a cystic component of a nodule with mixed composition, are a sonographic sign of benign disease [115]. However, when multiple calcifications and punctate echogenic foci with a comet-tail artefact are present in a hypoechoic solid nodule that can result in a high malignancy rate and PPV (77.8% and 96%, respectively) [283]. As such, CAD may misinterpret comet tail artefact as microcalcifications. This may account for the misdiagnosis of benign nodules interpreted as having mixed calcifications in our study. The TIRADS category 4 cut-off criteria, likewise, possibly contributed to the misdiagnosis findings because it denotes intermediate suspicion in some TIRADS, and this results in diagnostic challenges even with human assessment. These misdiagnoses affirm the necessity of clinical correlation in the accurate diagnosis of thyroid nodules even with the complementary use of thyroid ultrasound CAD.

3.5.2 Comparative analysis of CAD and computer-assisted subjective interpretation

The study demonstrated that for matched pairs of TIRADS, although the two approaches have comparable diagnostic performance, computer-assisted subjective interpretation using KSThR yielded a higher overall diagnostic accuracy than computer-aided diagnosis.

3.5.2.1 Sonographic feature ratings and rater agreement between the CAD and computer-assisted approaches

This study's findings showed generally high rank correlation associations between the computer-assisted subjective raters for echogenicity, calcifications, and composition. The association of margin ratings was negligible; implying that the two computer-assisted raters mainly varied in this sonographic rating. Because margin characteristics are among sonographic features that are highly predictive of malignancy, these differences likely affected the online calculator final malignancy-risk computation for AACE, EU and KSThR TIRADS. For the CAD vs either subjective rater, there was moderate association mainly for echogenicity and shape. While the rater agreement was mostly moderate between R_1 and R_2, the comparable sensitivities and specificities reflect how the computer-assisted scoring approach accounts for diverse rating criteria in determining a risk category.

CAD's reliance on textural and statistical feature analysis based on supervised machine learning may explain the moderate correlation association, moderate inter-rater agreement, and comparable sensitivities between CAD and each of the raters [284, 285]. While individual sonographic ratings may differ, CAD diagnosis outputs depend on the detected sonographic features within an automated or selected ROI. The sonographic features that are detected within an ROI depend on the algorithm that was used for training the CAD in malignancy risk stratification of different images. Hence, although CAD can interpret the same sonographic features that are rated by a subjective interpreter on the risk calculator model, CAD is affected

more by image quality thereby contributing to the increased sensitivity in misinterpreted suspicious features with CAD. Contrarily, an experienced human assessor may still be able to accurately interpret an image with artefacts that CAD is sensitive to.

3.5.2.2 Comparative analysis of diagnostic performance outcomes

The present study showed that the two raters independently using an online-based risk calculator had similar sensitivity and good diagnostic accuracy based on AUROC, with higher specificity across all TIRADS than the CAD. However, statistically significant differences in specificity were only observed using KSThR and AACE. Complementary to sensitivity and specificity outcomes, the DOR, PLR outcomes may aid the choice of TIRADS and approach to consider for clinical adoption as they are not prevalence-dependent [286, 287]. For all 4 TIRADS, the PLR were generally higher for the computer-assisted subjective raters than the CAD, and the DOR was highest with R_1 using any of the TIRADS (>9). For EU and KSThR, there was comparable sensitivity between CAD and both raters; however, there were statistically significant differences between R_1 and R_2. This implies that CAD systems are potentially an objective second opinion resource where there is ambiguity in subjective outputs. However, with the option of simultaneous output for multiple TIRADS, automated web-based risk systems may overcome challenges with subjective ambiguity and the bias of commercially-available CAD towards high sensitivity and low specificity. Although, deep learning-based CAD approaches have been suggested to result in improved accuracy and specificity; current studies on the commercially-available deep learning-based S-Detect 2 (Samsung Medison Co. Ltd., Seoul, South Korea), still show low specificity [237, 243-245]. AmCAD-UT used in the present study is also deep-learning-based for automated ROI selection, yet it similarly resulted in lower specificities than the computer-assisted approach.

Previous studies have established the comparable sensitivity but low specificity of commercial and non-commercial CAD systems to that of experienced clinicians [234, 242, 245, 251]. Only a few studies have shown higher specificity with CAD in comparison with subjective assessors of variable experience [248, 249]. A recent multi-centre study on the CAD-based on KSThR yielded a good AUROC (0.75) with the highest sensitivity (90.5%) but lowest specificity (49.6%) than that of the radiologists regardless of their experience [246]. However, in the present study, the KSThR TIRADS had the highest AUROC with R_1 and R_2 (0.76 and 0.74, respectively) with the highest sensitivity achieved by computer-assisted rater R_1 (SEN: 90.3%; SPEC: 51%) whereas CAD had a lower AUROC and specificity, but comparable sensitivity (0.67, 46% and 83.9%, respectively). The multi-centre study suggested that CAD KSThR be reserved for large cancer screening with subjective assistance supplemented by another TIRADS to increase specificity. However, this present study's findings, more so, the lowest NLR of 0.19 (0.09; 0.42) by R_1 using KSThR, suggest that the computer-assisted subjective rating approach would be better than CAD for ruling out disease. While ATA demonstrated good nodule discriminating ability (AUROC $\geq$ 0.7) overall for both approaches, this was achieved with 96 nodules due to a high rate of non-specified nodules using CAD (30.2% vs R_1 = 16% vs R_2 = 12.3%). Computer-assisted subjective rating with AACE specified all nodules whereas CAD failed to specify 29.6% thereby suggesting the superior efficiency of the risk calculator model for this TIRADS than the CAD.

3.6 Limitations

There were limitations in this present study. Due to the retrospective nature of the study, the selection of patients' images with FNAC and/or histopathology results, selection bias cannot be excluded. Furthermore, some histopathological diagnosis data of some nodules were not available which prevented the analysis of pathological factors in the validation aspect of the study. Another limitation was the small sample size in both aspects of the study whereby the malignant nodules were predominantly papillary thyroid cancer. In addition, the prevalence of malignancy within this study (38-40%) may not be reflective of the actual prevalence in the general population. The number of raters in this study was few to generalize the results to ultrasound users of variable experience. Because of the limited evaluation of the diagnostic performance of CAD using multiple TIRADS and readers, future studies are warranted to verify the influence of the CAD user experience based on different TIRADS and settings. Furthermore, there remains a need to explore the value of real-time subjectively-assisted CAD in comparison with retrospective automated CAD analysis and multiple computer-assisted raters of diverse experience levels.

The shortcomings in the present study limit the generalization of the study findings to the general population and other thyroid cancers. Therefore, larger standardized multi-rater and multi-centre prospective studies with a diverse representation of thyroid cancers are warranted to assess the validity and generalizability of the findings.

3.7 Conclusions

This study validated the diagnostic performance of AmCAD-UT to best for all 6 TIRADS under investigation, at the default setting. We established that the default setting was best for maximal sensitivity with all TIRADS, with EU-TIRADS having the highest sensitivity. However, there is likely a potential to improve specificity with minimal compromise to the sensitivity at a hyperechoic foci detection setting of 3.5 with other settings maintained at median values. This setting may require exploration when specificity must be maximized to limit unnecessary biopsy rates.

The computer-assisted subjective assessment approach had comparable diagnostic performance to that of the CAD approach for all the 4 TIRADS. However, when using KSThR, the lower specificity of CAD outweighs its high sensitivity thereby resulting in a lower diagnostic accuracy than the computer-assisted subjective approach. The best DOR of almost 10 compared to about 5 for CAD for rating all nodules, imply that computer-assisted subjective rating is superior to CAD when using KSThR or EU in the detection of mainly papillary thyroid carcinomas. However, both approaches have the potential for clinical diagnostic workflow adoption in thyroid ultrasound imaging for screening purposes due to comparable high sensitivities and NPVs. Yet, where dual use of both approaches can be implemented based on the TIRADS that specify all nodules, computer-assisted subjective interpretation using KSThR may be the best choice for rule out purposes. EU with either approach may suffice for rule-in purposes whereas the use of ATA and AACE may be best with computer-assisted subjective interpretation since CAD has higher rates of non-specified nodules.

Chapter 4

Study Two: The Diagnostic Efficiency of AngioPLUS Microvascularity Doppler Detection in Combination with EU TIRADS for Distinguishing Benign and Malignant Thyroid Nodules

4.1 Introduction

TIRADS guidelines based on qualitative and quantitative scoring of suspicious features have emerged to overcome the drawbacks of sole sonographic features in malignancy risk prediction. However, different TIRADS guidelines vary in their malignancy-risk estimation criterion of suspicious grey scale features, and in some TIRADS such as EU and K TIRADS, nodule vascularity is not incorporated in the risk stratification categories [135, 136].

Increased vascularity or microvascularization are anticipated consequences of abnormal angiogenesis during carcinogenesis as cancer cells invade the nodule areas that are deficient in blood vessels [16, 17]. Conventional colour and power Doppler ultrasound have been commonly used in thyroid nodular vascularity assessment to complement grey scale ultrasound findings. Marked central vascularity, intranodular vascularity and/or increased chaotic intranodular central vascularity on colour and power Doppler ultrasound has been suggested to be associated with suspicion for malignancy in some studies while peripheral vascularity is linked with benignity [18, 19, 126, 157]. However, this assertion remains contested due to variable findings and some studies have suggested that vascularity patterns on colour and power Doppler ultrasound have little value in malignancy prediction even when combined with grey scale ultrasound features [132, 159, 164]. The poor diagnostic performance of conventional Doppler methods for malignancy prediction can also be attributed to the poor sensitivity of these methods in microvascularity pattern assessment.

Microvascular ultrasound imaging techniques such as superb microvascular imaging (SMI) and Angio Planewave Utrasensitive (AngioPLUS) imaging have emerged in recent years. These methods outperform colour and power Doppler ultrasound (also known as colour flow imaging, CFI, and power Doppler imaging, PDI) in depicting microvascular flow, small blood vessel branching and low blood flow velocity [21, 172]. Studies conducted with SMI on thyroid nodules have suggested that it has superior accuracy than CFI and PDI at levels comparable to those of CEUS in malignancy prediction [175, 176]. On the other hand, the diagnostic utility of AngioPLUS in thyroid nodule vascularity assessment for malignancy prediction lacks exploration. The few studies that have evaluated the grey scale ultrasound plus SMI combinations for thyroid malignancy prediction have not focused on specific TIRADS. Current studies on SMI have only focused on qualitative microvascularity assessment and have not evaluated any value of the technique in thyroid nodules with equivocal cytology. Microvascularity assessment for malignancy risk stratification in these nodules may be of more value in these nodules which are suggested to generally have increased vascularity than other nodules [180].

This current study sought to determine the diagnostic value of AngioPLUS in discriminating benign and malignant thyroid nodules using a qualitative approach and quantitative regional vascularity ratio analysis using a locally established image processing computer algorithm. The diagnostic performance of the best vascularity assessment approach based on conventional Doppler ultrasound modes and/or AngioPLUS modes in combination with EU TIRADS was evaluated for all nodules and the subgroup of nodules with equivocal cytology. The complementary and combined use of AngioPLUS with EU TIRADS has the potential to improve overall diagnostic accuracy and limit unnecessary FNAC, particularly in thyroid nodules with equivocal FNAC results.

4.2 Materials and methods

4.2.1 Study type

This was a prospective analytical observational study that received ethical approval from the Institutional Human Research Ethics sub-committee before it was conducted. Consecutive case analysis and non-probability sampling were applied. Cross-sectional cohorts of patients with thyroid nodules and/or suspicion of thyroid cancer were purposively recruited at the Prince of Wales Hospital Department of Surgery and its affiliates from 15 May 2019 to 31 August 2021. Informed consent was sought from the patients before the data collection procedures.

4.2.2 Data collection procedures

4.2.2.1 Inclusion and exclusion criteria

The inclusion criteria in this study were all consenting adult patients (≥18 years old) who had thyroid nodular disease or suspicion of thyroid cancer and were scheduled for FNAC and/or subsequent thyroidectomy. Nodules that were between ≥ 5mm were included in the study. For patients with multiple thyroid nodules, either the nodule with the most suspicious sonographic features (hypoechoic, microcalcifications, irregular margins, tall-than-wide, etc.), or if there were no obvious suspicious features then the largest nodules on each lobe or the one/s for which biopsy and/or surgery were recommended, were included in the study. The exclusion criteria were patients < 18 years old, patients who had a thyroidectomy, did not have a conclusive diagnosis as determined from either cytology results, histopathology results or both, and multinodular goitres without clearly isolated nodules. Patients with only completely cystic nodules and nodules that were too large for the footprint of the transducer were excluded from the study as they would affect the elastography output [183].

4.2.2.2 Ultrasound imaging procedures

A sole investigator (N.C.) conducted the thyroid ultrasound imaging of all patients using the same Aixplorer ultrasound machine (Supersonic Imagine, Aix-en-Provence, France) equipped with a 7-10 MHz linear transducer, AngioPLUS and SWE. The ultrasound machine settings were standardized during a pilot study, and the same ultrasound scanning settings for the thyroid study were maintained throughout the study to ensure consistency. For the Doppler ultrasound modes, the standard settings were set with a medium wall filter, at the lowest pulse repetition frequency where no aliasing was encountered and the highest colour gain without signal noise. The resultant settings used in the study were a velocity scale of 10cm/s with a colour map of 5 for CFI and PDI and a velocity scale of 4cm/s with a colour map of 4 for the AngioPLUS modes.

The investigator followed standard ultrasound protocols to conduct the thyroid scans. Each patient lay in the supine position with minimal extension of the neck, and coupling gel was applied. With the face turned away from the side of interest, the transducer was placed on the exposed side of the neck area. A minimum of 3 transverse and 3 longitudinal images of each target thyroid nodule in grey scale, colour Doppler, power Doppler, both colour and power Doppler with AngioPLUS, and SWE modes were acquired. The images of each target nodule were acquired by scanning through the superior, mid and inferior pole of the thyroid gland and capturing the nodule where the suspicious ultrasound feature/s were best visualised, and the point of maximal transverse and anteroposterior diameter for size measurements. The longitudinal images were scanned from the lateral, mid and medial regions and the images were acquired at points that fully demonstrated the nodule and at the longest axis view for measurements. In the Doppler ultrasound mode, images were obtained at the scan planes where the nodule showed abundant vascularity and stable Doppler ultrasound signals. In SWE mode,

images were acquired at similar regions where the nodule demonstrated visible and stable SWE signals.

4.2.2.3 Grey-scale ultrasound feature assessment

The same subjective and CAD approaches that were applied in section 3.3.2 were used for this assessment; however, one rater (N.C.) solely conducted the subjective analyses in this study. The analysis outputs were compared between the CAD and subjective approaches of EU TIRADS which were concluded as optimal for rule in purposes in Chapter 3. The common features that are indicative of high malignancy suspicion in EU TIRADS like most RSS, are irregular margins, marked hypoechogenicity, taller than wide shape, and microcalcifications [288]. However, for the classification of high-risk categories unlike other RSS e.g. KSThR, that consider the composition and echogenicity of a nodule (solid and hypoechoic) in addition to the presence of any of the suspicious features, EU TIRADS requires the presence of any of the common suspicious features or marked hypoechogenicity in a solid nodule [135, 136]. The categorisation criteria for the EU TIRADS are shown in Figure 4:1. In this study, EU cut-off category 5 was used for malignancy risk stratification for sole grey scale assessment as was previously determined in Chapter 3. We hypothesized that an additional suspicious feature has more potential to improve overall diagnostic accuracy from the level of indeterminate ultrasound suspicion rather than in category 5 which already represents high suspicion. Therefore, EU category 4 was used as the cut off point for the combined diagnostic performance assessment.

4.2.2.4 Doppler ultrasound feature assessment

The vascularity features of the nodules were assessed using CFI and PDI and both modes with AngioPLUS based on a qualitative and quantitative approach.

4.2.2.4.1 Qualitative vascularity assessment

Each nodule was subjectively evaluated for vascularity for all 4 Doppler modes (i.e. 1. CFI, 2. AngioPLUS+CFI (ACFI), 3. PDI, 4. AngioPLUS+PDI (APDI). The subjective assessment was based on the qualitative grading criteria adapted from Chammas *et al* [161]: Category I = exclusively peripheral vascularity; Category II = predominantly peripheral vascularity; Category III = predominant central vascularity and, Category IV = exclusively central blood flow. Nodules with categories I and II features were considered most likely benign while those falling in categories III and IV were considered most likely malignant (*Figure 4:2)*. The Doppler mode that resulted in the best diagnostic performance using this approach was then compared to the grey scale assessment and used in the combined assessment of diagnostic performance.

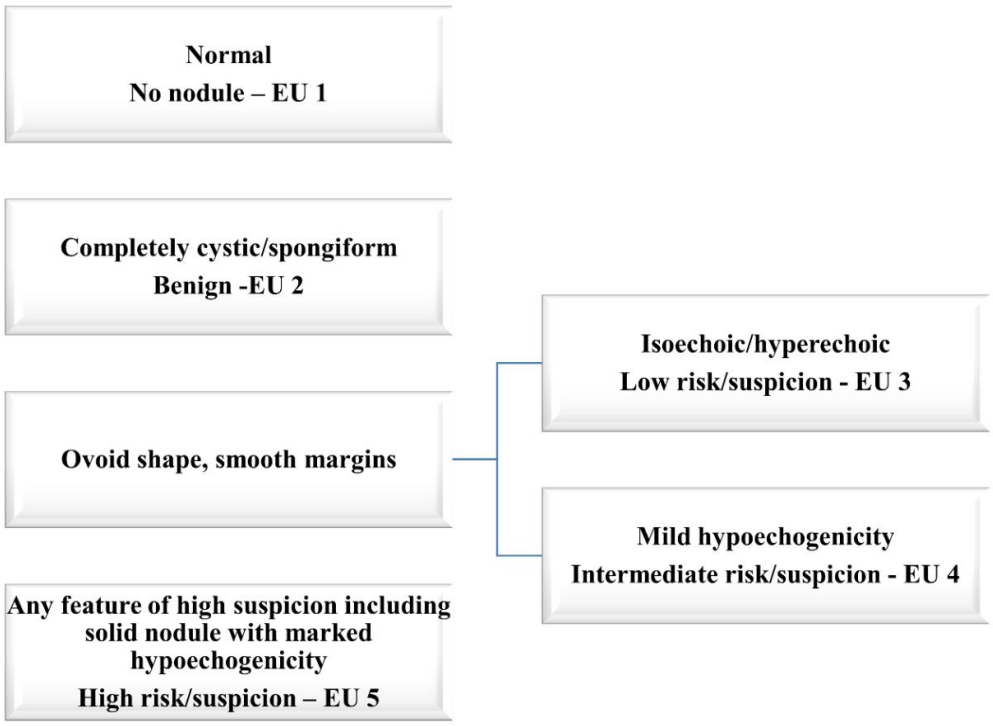


Figure 4:1 Predictive ultrasound features for the different EU TIRADS categories

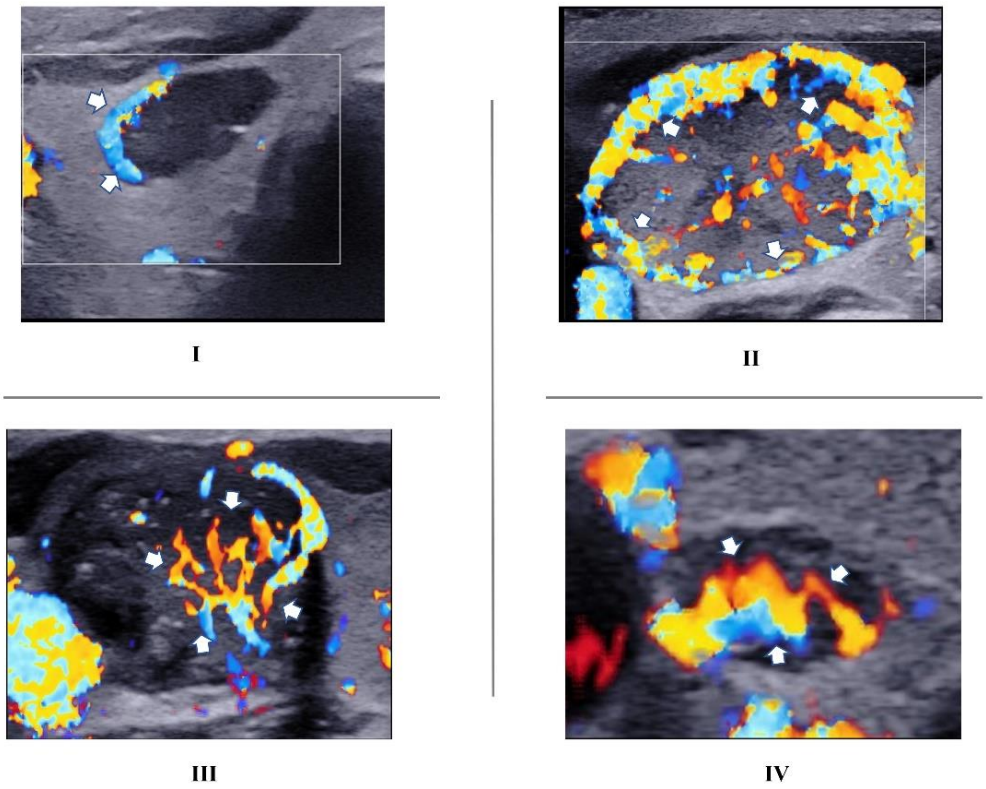


Figure 4:2 CFI Images demonstrating the adopted qualitative vascularity grading.

I: Exclusive peripheral vascularity (arrows) in a nodule that had a conclusive benign histopathology diagnosis in a 61-year-old male patient. II: Predominant peripheral vascularity demonstrated as more abundant vascularity in the outer regions (arrows) than the central regions in a histopathologically-benign nodule in a 53-year-old female patient. III: Predominant vascularity in the central portion of a nodule (arrows) in a 44-year-old female who was diagnosed with PTC. IV: Exclusive marked central flow (arrows) in a nodule diagnosed as PTC in a 51-year-old female.

4.2.2.4.2 Quantitative vascularity assessment

An image processing computer algorithm that segments regions of central and peripheral vascularity on CDUS and PDUS images to determine regional vascularization based on offsetting, which was previously established by our research group was used for the quantitative evaluation of thyroid nodule vascularity [23, 168, 231]. The sets of three ultrasound images of the thyroid nodules acquired for each Doppler mode in both transverse and longitudinal scan planes were observed, documented, and saved for offline analysis in MATLAB (version 9.4.0.813654 R2018a; The Math Works, Natick, MA, USA). Microsoft Paint was used to manually outline regions of interest (ROI) and the images were saved in TIFF format and processed further in MATLAB. In MATLAB, the local algorithm extracted the primary ROI (i.e. the thyroid nodule) from the outlined area and quantitatively analysed the vascularity of the thyroid nodules based on the number of colour pixels in the primary ROI over the total number of pixels of the primary ROI. Afterwards, using the algorithm, an offsetting method was used to extract the secondary ROI from the primary ROI. This method adjusts the area of a shape as a percentage of the maximum diameter of the primary ROI without contour changes or shape distortion of the ROI. At a given offset level e.g., 10%, the secondary ROI which represents the central region of the nodule that was extracted from the primary ROI. The remaining outer segment of the primary ROI represents the peripheral region of the nodule (*Figure 4:3).* With the central and peripheral areas of the nodule segmented, the computer algorithm computes the total number of pixels within the segmented areas. The colour pixels coded by the Doppler ultrasound are extracted and computed by the algorithm through the elimination of grey scale pixels. The vascular indices (VI) of peripheral and central thyroid nodule regions are the percentages of the number of colour pixels to the total number of pixels within each segmented area. The overall VI of the thyroid nodule is calculated by the total number of pixels within the primary ROI and the number of colour pixels in the image

excluding the grey scale pixels. Figure 4:4 illustrates the algorithm's segmentation of central and peripheral vascularity of a thyroid nodule at the 10% offset.

In this study, pilot analysis of the vascularity quantification assessment on a set of 40 nodules in the transverse plane, longitudinal plane and the combination of both planes showed no statistically significant differences in the computed mean total, central and peripheral VIs between the 3 approaches. Hence, the average of the three sets of readings for each of CFI, PDI, ACFI, and APDI in the transverse plane was used to determine the total, central and peripheral VIs. The previous study by our group established that the 22% offset yielded the optimal diagnostic performance for the regional segmentation of central and peripheral vascularity based on CDI [23]. In this present study, ratio analysis of the central and peripheral VIs was applied whereby a ratio vascularity index (RVI) of peripheral VI to central VI > 1 denoted predominant peripheral vascularity and ≤ 1 denoted predominant central vascularity. Different offsets were employed incrementally to determine the offset that resulted in the optimal diagnostic performance using the RVI method for all 4 Doppler modes. The Doppler mode with the best diagnostic performance was compared to the grey scale and qualitative approaches and assessed in combination with the grey scale assessment.

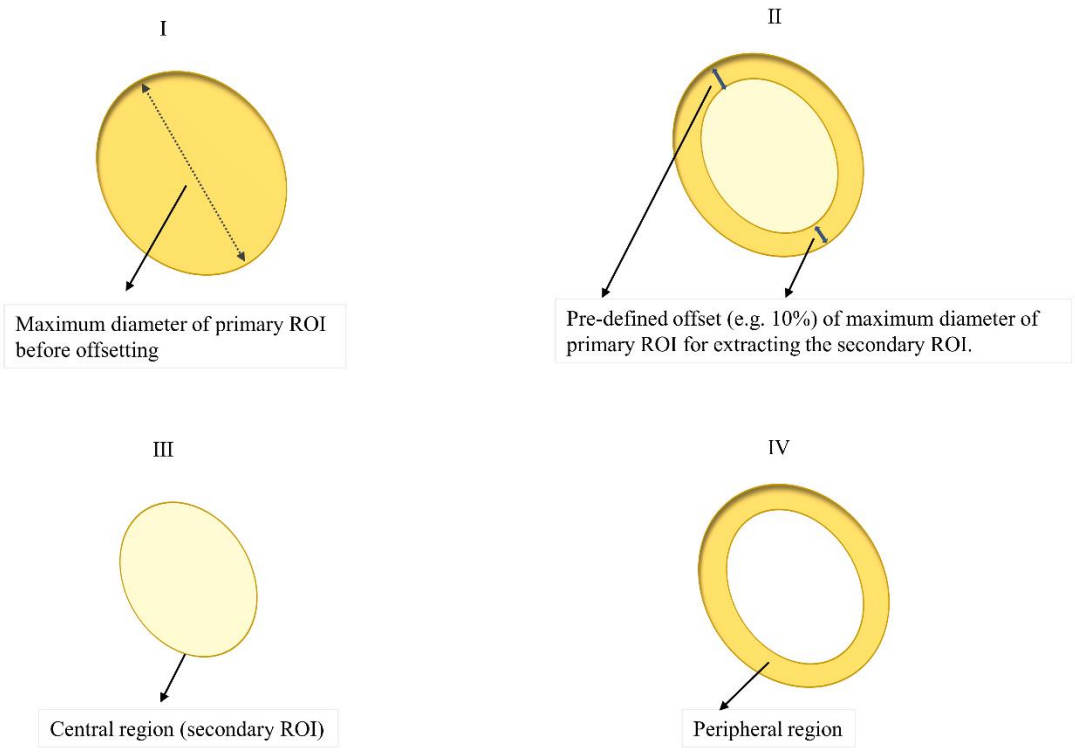


Figure 4:3 A schematic representation of the offsetting principle for extracting the secondary ROI from the primary ROI.

I: The primary ROI that represents the thyroid nodule. II: Application of a pre-defined offset level as a percentage of the maximum diameter to extract the secondary region. III: The extracted secondary ROI which represents the central region of the nodule. IV: The peripheral region or outer region of the nodule acquired from subtracting the secondary ROI from the primary ROI.

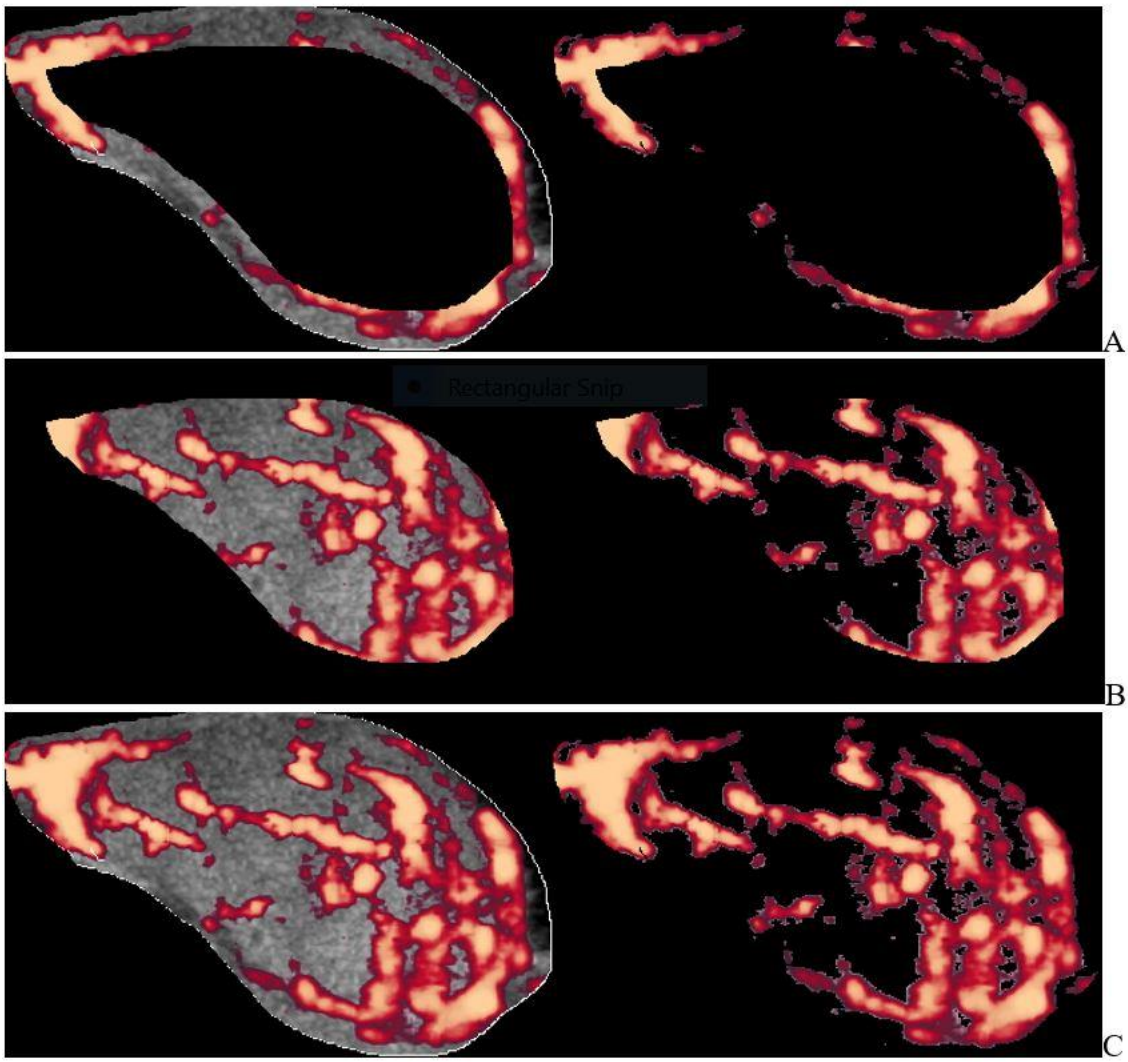


Figure 4:4: Segmentation of thyroid nodule vascularity into A) peripheral and B) central regions. C shows the primary ROI from which the overall vascularity is calculated.

The VI for the different segments is calculated by the algorithm as the percentage between the total number of pixels (left) and the number of colour pixels without the grey scale pixels (right), within the segmented areas.

4.2.3 Data analysis and statistical analysis

The data analysis from this prospective study involved the assessment of all vascular nodules and the analysis of vascular nodules with equivocal FNAC results alone to determine the diagnostic value of the approaches. Nodules that had FNAC classifications 3 and 4 (AUS/SFN/FN/FLUS) were considered equivocal.

Continuous data were classified as means +/- SD whereas categorical and/or nominal data were expressed as frequencies and percentages. The Shapiro Wilk test was used to check the normality of the data. The Chi-square test was used to compare differences in nodule classification data and peripheral and central vascularity index ratios of benign and malignant nodules. The paired samples T-test was used for testing statistically significant differences in the mean central and peripheral vascularity index quantification between the CFI and ACFI, as well as PDI and APDI Doppler modes. The Cohen's kappa statistic (κ) complemented by the proportion agreement test was used for the intra-rater reliability assessment of qualitative grading of nodule vascularity using different Doppler modes. The SEN, SPEC, PPV, NPV and AUROCs and their corresponding 95% confidence intervals were calculated with reference to the final cytology or pathology results. The ROCs were generated to obtain the AUROCs. For the combined assessment of EU TIRADS and either qualitative or quantitative vascularity, a nodule was predicted to be malignant if it was category 4 or 5 and met the malignancy cut-off criteria for the vascularity approaches. The McNemar and Cochran Q's tests were used for the comparative analysis of sensitivity and specificity whereas the z-test was used to compare the different AUROCs. Multi-comparison testing was not applied to employ a more conservative approach and limit the false-negative rate (type 2 error) as is desirable in diagnostic accuracy testing [289]. The tests were two-sided and $p < 0.05$ denoted statistical significance.

4.3 Results

4.3.1 Demographic data

There was a total of 94 nodules (30 malignant; 64 benign) from 92 patients (78 females; 14 males). The mean age of all patients was 53 ± 12.8 years (range: 21 to 75). There was a statistically significant difference in the mean age between male (60.7 ± 9.5; range 44 to 71) and female (51.7 ± 12.9; range 21 to 75) patients ($p = 0.01$). However, there was no statistically significant difference in age between patients with malignant and benign thyroid nodules (53.5 ± 12.7; range: 31 to 74, and 52.9 ± 13; range: 21 to 75, respectively, $p = 0.84$). The most common histopathology diagnosis of the malignant nodules was PTC (n=26) while the remaining nodules were classified as FTC (n=3) and NIFTP (n=1). There were 40 nodules (31 benign; 9 malignant) with equivocal cytology. Two of the malignant nodules with equivocal cytology were FTCs while the rest were PTCs.

4.3.2 Grey scale assessment findings

The grey scale ultrasound findings of the nodules based on the subjective and CAD approaches were analysed. The results showed comparable overall diagnostic performance for both the CAD and subjective approach for the EU TIRADS as was previously established in Chapter 3 (*Table 4.1*). For the analysis of all nodules and the subgroup of nodules with equivocal cytology, the sensitivity and specificity were comparable for the EU- CAD/ Subjective method pairs, however, the diagnostic performance outcomes were slightly higher with the subjective approach. The overall diagnostic performance based on the AUROC was higher with EU using subjective assessment for all nodules and the subgroup of cytologically equivocal nodules, respectively (0.67 and 0.62).

Table 4.1: Diagnostic performance assessment for grey scale ultrasound assessment in vascular nodules

Nodules	**Diagnostic Performance Measures**	**EU_Sub**	**EU_CAD**
All (N= 94)	SEN (%)	83.3	76.7
	SPEC (%)	50.0	45.3
	NPV (%)	86.5	80.6
	PPV (%)	43.9	39.7
	AUROC	0.67	0.61
Equivocal (N=40)	SEN (%)	88.9	77.8
	SPEC (%)	38.7	41.9
	NPV (%)	92.3	86.7
	PPV (%)	29.6	28.0
	AUROC	0.62	0.61

Sub = Subjective Analysis, CAD = Computer-aided diagnosis

4.3.3 Qualitative and quantitative vascularity assessment using different Doppler ultrasound modes

The vascularity of the thyroid nodules based on CFI, ACFI, PDI, and APDI was analysed using qualitative and quantitative approaches.

4.3.3.1 Qualitative vascularity assessment results

The intra-observer reliability of using the qualitative method in grading thyroid nodule vascularity was assessed and demonstrated moderately substantial reliability for all the Doppler modes (*Table 4.2*). The diagnostic performance analysis for the qualitative vascularity assessment of all nodules and the subgroup of nodules with equivocal cytology showed that the sensitivity was lower with CFI and PDI modes than their respective AngioPLUS modes whereas the specificity was higher. Table 4.3 shows the diagnostic performance outcomes. For the classification of all nodules, the AngioPLUS modes had higher overall diagnostic performance based on the AUROC, with APDI showing statistically significant differences from the PDI mode (0.82 vs 0.71, $p = 0.04$). There was excellent sensitivity, good specificity and very good AUROC in the assessment of nodules with equivocal cytology with APDI (100%, 77.4%, and 0.89, respectively).

4.3.3.2 Quantitative vascularity assessment results

The quantitative assessment of thyroid nodule vascularity was analysed based on regional segmentation of peripheral and central regions. At the different offsets, the mean vascularity indices were significantly higher with both AngioPLUS modes than with the sole CFI and PDI modes, in quantifying peripheral and central vascularity in benign and malignant nodules, $p < 0.05$ (*Figure 4:5 and Figure 4:6*). The analysis of regional vascularity computed as a ratio (RVI) showed that for all offsets and all Doppler modes, benign nodules tended to have predominantly peripheral vascularity (RVI >1) and the results were statistically significant

(*Table 4.4*, $p < 0.01$). There were no statistically significant differences between predominantly peripheral and predominantly central vascularity in malignant nodules (*Table 4.4*, $p > 0.05$). In the subgroup of nodules with equivocal cytology, most benign nodules also presented with predominantly peripheral vascularity than central vascularity with statistically significant differences ($p < 0.01$) for all Doppler modes at ≥ 18% offset, except PDI.

The diagnostic performance of the RVI > 1 method to denote predominantly peripheral vascularity was evaluated for all the Doppler modes at different offsets (*Table 4.5*). For all nodules, the results showed comparable low sensitivity for each Doppler mode at the different offsets. The RVI method generally had higher specificity than sensitivity, however, of all the Doppler modes PDI had poor AUROC overall. The highest sensitivity of 66.7% was achieved with the APDI at the 22% offset whereas the highest specificity of 84.4% was achieved with CFI at the 18% offset. The highest discriminatory ability shown by an AUROC of 0.71 was achieved with CFI at the 20% offset. For the subgroup of nodules with equivocal cytology, the 22% and 25% offsets with APDI resulted in the highest sensitivity and AUROC overall. (SEN: 77.8%, SPEC: 74.2% and AUROC: 0.76). However, only the sensitivity was statistically significant from that of all other Doppler modes at different offsets ($p < 0.05$), while the AUROC was comparable to that of CFI. RVI outcomes with CFI at the 20% offset for all nodules and APDI at the 22% offset for equivocal nodules were therefore analysed further in combination with EU TIRADS (*Section 4.3.4 and Table 4.6*).

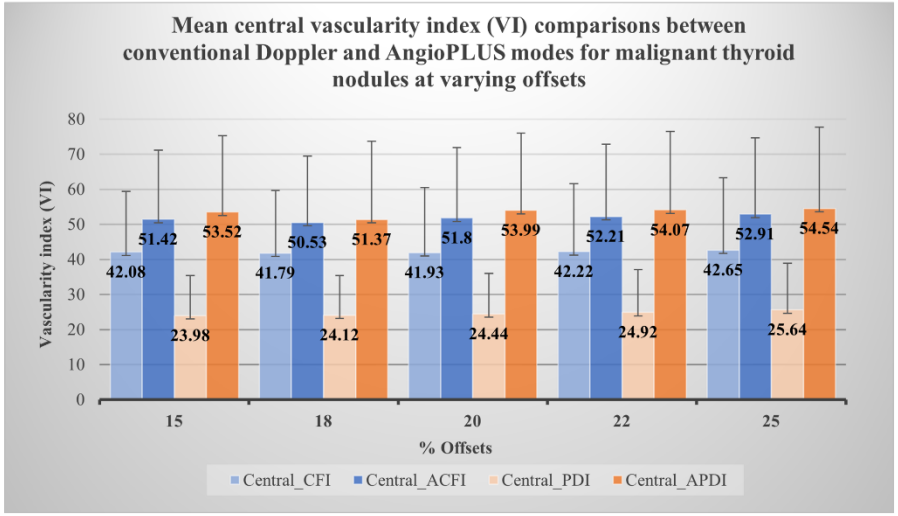
Table 4.2: Intra-rater reliability (IRR) testing for the qualitative grading of thyroid nodule vascularity using different Doppler modes

Vascularity Mode	Kappa rate of agreement (κ)	Proportions agreement (%)
CFI	0.75	80
PDI	0.64	77
ACFI	0.76	80
APDI	0.71	77

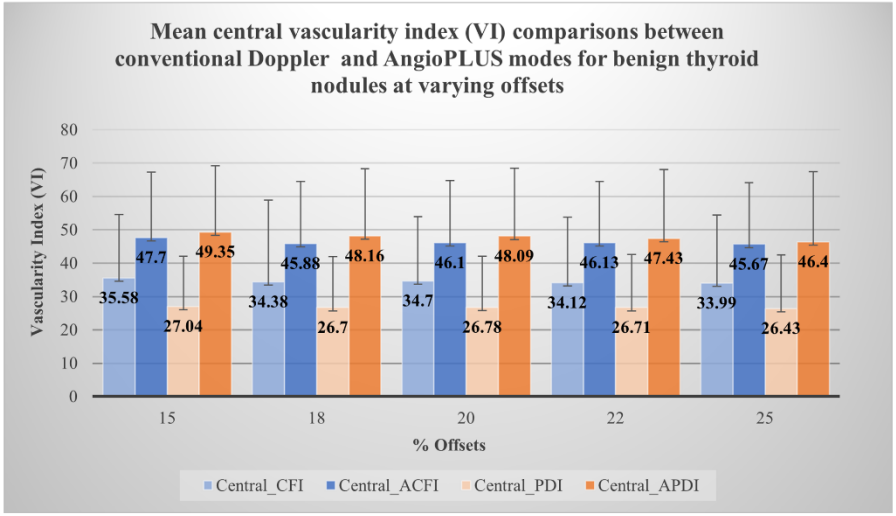
Table 4.3: Diagnostic performance of qualitative vascularity assessment of different Doppler modes in thyroid nodule differentiation

Nodules	**Diagnostic Performance Measures**	**CFI_Qual**	**ACFI_Qual**	**PDI_Qual**	**APDI_Qual**
All (N = 94)	SEN (%)	53.3*	80.0	46.7*	83.3
	SPEC (%)	92.2	82.8	95.3	81.3
	NPV (%)	80.8	89.8	79.2	91.2
	PPV (%)	76.2	68.6	82.4	67.6
	AUROC	0.73	0.81	0.71*	0.82
Equivocal (N = 40)	SEN (%)	66.7*	88.9	66.7*	100
	SPEC (%)	90.3	80.6	93.5	77.4
	NPV (%)	90.3	96.2	90.6	100
	PPV (%)	66.7	57.1	75	56.3
	AUROC	0.79	0.85	0.80	0.89

*– * $p < 0.05$ relative to APDI and/ or ACFI, Qual: qualitative assessment*



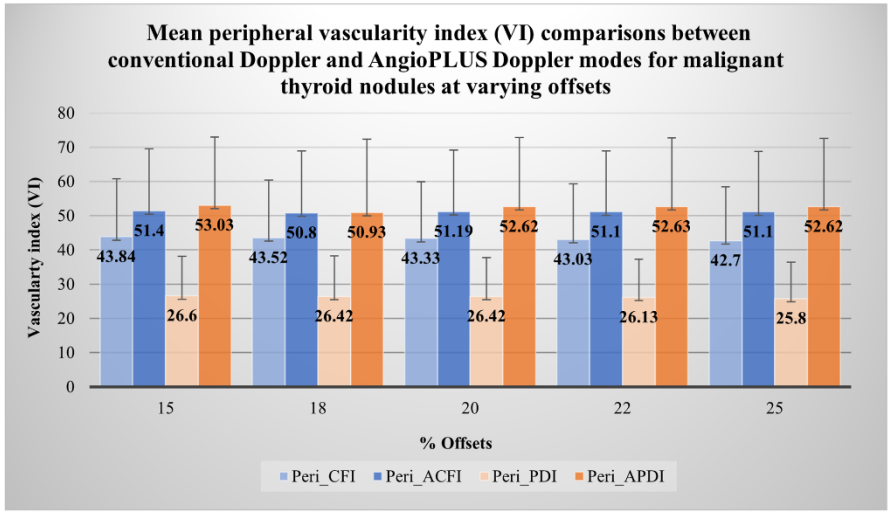
A



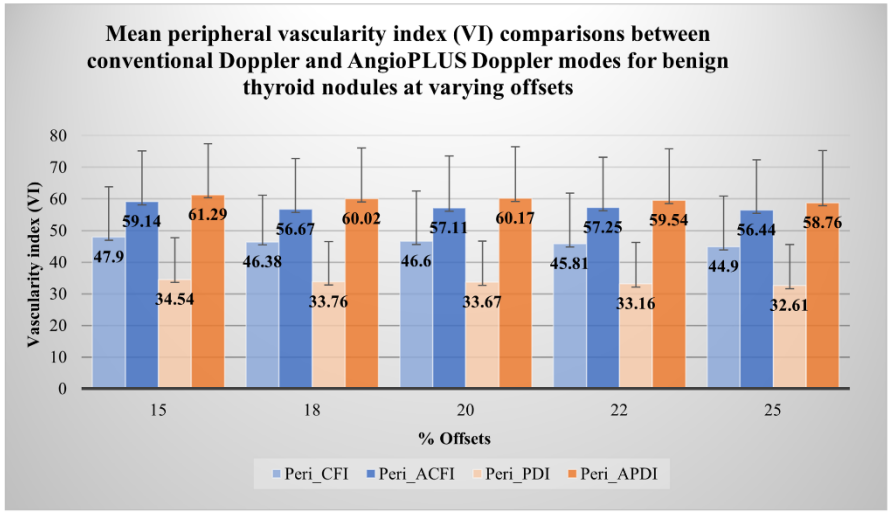
B

Figure 4:5: Graphical representation of differences in mean VIs for central vascularity between CFI vs ACFI, and PDI vs APDI modes

The mean VIs were significantly higher with AngioPLUS modes than conventional colour and power Doppler modes at all offsets for A) malignant, and B) benign nodules.



A



B

Figure 4:6: Graphical representation of differences in mean VIs for peripheral vascularity between CFI vs ACFI, and PDI vs APDI modes.

The mean VIs were significantly higher with AngioPLUS modes than conventional modes at all offsets for A) malignant, and B) benign nodules.

Table 4.4: Classification of predominantly peripheral (RVI > 1) and predominantly central vascularity (RVI ≤1) at different offsets based on diagnosis for all nodules and equivocal nodules

		ALL (BEN= 64, MLG=30)								EQUIVOCAL (BEN= 31, MLG=9)							
Offset level	**Nodule Status**	**CFI**		**PDI**		**ACFI**		**APDI**		**CFI**		**PDI**		**ACFI**		**APDI**	
		RVI >1	**RVI ≤1**	**RVI >1**	**RVI ≤1**	**RVI >1**	**RVI ≤1**	**RVI >1**	**RVI ≤1**	**RVI >1**	**RVI ≤1**	**RVI >1**	**RVI ≤1**	**RVI >1**	**RVI ≤1**	**RVI >1**	**RVI ≤1**
15%	BEN	53**	11	44**	20	45***	19	47***	17	24**	7	20	11	19	12	24**	7
	MLG	15	15	17	13	12	18	12	18	3	6	6	3	4	5	4	5
18%	BEN	54***	10	43**	21	47***	17	49***	15	26***	5	20	11	21**	10	25***	6
	MLG	13	17	17	13	13	17	12	18	3	6	6	3	4	5	4	5
20%	BEN	53***	11	44**	20	47***	17	49***	15	26***	5	19	12	21**	10	25***	6
	MLG	12	18	17	13	12	18	11	19	3	6	6	3	4	5	3	6
22%	BEN	52***	12	42**	22	47***	17	52***	12	26***	5	19	12	21**	10	23***	8
	MLG	13	17	16	14	12	18	10	20	3	6	6	3	4	5	2	7
25%	BEN	49***	15	41**	23	46***	18	47***	17	24**	7	19	12	21**	10	23***	8
	MLG	13	17	16	14	11	19	11	19	3	6	6	3	4	5	2	7

*BEN = benign; MLG = malignant; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$*

Table 4.5: Diagnostic performance of the RVI > 1 to indicate predominant peripheral vascularity to differentiate benign and malignant nodules at different offsets

		All nodules (N = 94)					Equivocal nodules (N = 40)				
Vascularity Mode	**Diagnostic Performance Measures**	**15% Offset**	**18% Offset**	**20% Offset**	**22% Offset**	**25% Offset**	**15% Offset**	**18% Offset**	**20% Offset**	**22% Offset**	**25% Offset**
CFI	SEN	50.0	56.7	60.0	56.7	56.7	66.7	66.7	66.7	66.7	66.7
	SPEC	82.8	84.4	82.8	81.3	76.6	77.4	83.9	83.9	83.9	77.4
	AUROC	0.66‡	0.71‡‡‡	0.71‡‡‡	0.69‡‡	0.67‡	0.72‡‡	0.75‡‡	0.75‡‡	0.75‡‡	0.72‡‡
ACFI	SEN	60.0	56.7	60.0	60.0	63.3	55.6	55.6	55.6	55.6	55.6
	SPEC	70.3*	73.4	73.4	73.4	71.9	61.3	67.7	67.7	67.7	67.7
	AUROC	0.65	0.65	0.67‡	0.67	0.68‡	0.58	0.62	0.62	0.62	0.62
PDI	SEN	43.3	43.3	43.3	46.7	46.7	33.3+	33.3+	33.3+	33.3+	33.3+
	SPEC	68.8*	67.2**	65.6**	65.6**	64.1	64.5	64.5	61.3	61.3	61.3
	AUROC	0.56	0.55	0.55	0.56	0.55	0.49	0.49	0.47	0.47	0.47
APDI	SEN	60.0	60.0	63.3	66.7	63.3	55.6	55.6	66.7	77.8+	77.8+
	SPEC	73.4	76.6	76.6	73.4	73.4	77.4	80.6	80.6	74.2	74.2
	AUROC	0.67‡	0.68‡	0.70‡‡	0.70‡‡	0.68‡	0.67	0.68	0.74‡‡	0.76‡‡	0.76‡‡

** = p < 0.05 , ** = p < 0.01 with reference to CFI; + = p < 0.05 with reference to all other modes; ‡ = p < 0.05, ‡‡ = p < 0.01, ‡‡‡ = p < 0.001 with reference to PDI*

4.3.4 Diagnostic performance of vascularity assessment approaches when combined with grey scale assessment

The vascularity assessment approaches with the highest diagnostic performance were combined with grey scale assessment to ascertain the combined diagnostic performance. Table 4.6 shows these results. Based on the qualitative vascularity assessment, there was high nodule discrimination ability when APDI was combined with EU TIRADS for the assessment of all nodules and equivocal cytology nodule subgroup (AUROC: 0.81 and 0.83, respectively). The sensitivity of this combined approach in stratifying all nodules was insignificantly lower than that of sole EU TIRADS and sole APDI (76.7% vs 83.3, $p > 0.05$) whereas the specificity was significantly higher than that of sole EU TIRADS (84.4% vs 50%, $p < 0.05$). The combined approach demonstrated sensitivity that was comparable to the sole EU approach in discriminating equivocal nodules with a specificity that was significantly higher than that of sole EU TIRADS (SEN: 88.9% vs 88.9%; SPEC: 77.4% vs 38.7%, $p < 0.05$).

For the quantitative assessment using RVI, the combined approach of 20_CFI (CFI at the 20% offset) with EU TIRADS in stratifying all nodules demonstrated a significantly raised specificity with a reduction in sensitivity when compared to sole EU TIRADS (SPEC: 89.1% vs 50%, $p < 0.05$; SEN: 56.7% vs 83.3%, $p < 0.05$). Despite the higher specificity, the overall discriminatory ability of the combined approach was comparable to that of sole EU TIRADS and the sole 20_CFI RVI method (AUROC: 0.73 vs 0.67; 0.73 vs 0.71, $p > 0.05$, respectively). For equivocal nodules, the combined approach of 22_APDI (APDI at the 22% offset) and EU TIRADS improved the specificity of EU TIRADS with a reduction in sensitivity (SPEC: 38.7% to 74.2%, $p < 0.001$ and SEN: 88.9% to 66.7%, $p > 0.05$). The specificity of the combined approach remained similar to that of the sole RVI approach while the sensitivity decreased (SEN: 77.8% vs 66.7, $p > 0.05$; SPEC: 74.2% vs 74.2%).

Table 4.6: Diagnostic performance results of the combined grey scale and vascularity approaches

Nodules	**Diagnostic Performance Measures**	**GSU_TIRADS** **EU**	**Qualitative vascularity mode** **APDI**	**GSU + Qualitative vascularity** **EU + APDI**	**Quantitative vascularity mode** **20_CFI**	**GSU + Quantitative vascularity** **EU + 20_CFI**
All (N = 94)	SEN (%)	83.3 (65.3 ; 94.4)	83.3 (65.3 ; 94.4)	76.7 (57.7 ; 90.1)	60.0 (40.6 ; 77.3)$^{*\,\S}$	56.7 (37.4 ; 74.5)$^{*\,\S}$
	SPEC (%	50.0 (37.2 ; 62.8)	81.3 (69.5 ; 89.9)***	84.4 (73.1 ; 92.2)***	82.8 (73.3 ; 91.9)***	89.1 (78.8 ; 95.5)***
	PPV (%)	43.9 (30.7 ; 57.6)	67.6 (50.2 ; 82.0)	69.7 (51.3 ; 84.4)	62.1 (42.3 ; 79.3)	70.8 (48.9 ; 87.4)
	NPV (%)	86.5 (71.2 ; 95.5)	91.2 (80.7 ; 97.1)	88.5 (77.8 ; 95.3)	81.5 (70.0 ; 90.1)	81.4 (70.3 ; 89.7)
	AUROC	0.67 (0.57 ; 0.76)	0.82* (0.74 ; 0.91)	0.81* (0.72 ; 0.89)	0.71§ (0.61 ; 0.82)	0.73 (0.63 ; 0.83)
Equivocal (N = 40)		**GSU_TIRADS** **EU**	**Qualitative vascularity mode** **APDI**	**GSU + Qualitative vascularity** **EU+ APDI**	**Quantitative vascularity mode** **22_APDI**	**GSU + Quantitative vascularity** **EU+ 22_APDI**
	SEN (%)	88.9 (51.8 ; 99.7)	100 (66.4 ; 100)	88.9 (51.8 ; 99.7)	77.8 (40.0 ; 97.2)	66.7 (29.9 ; 92.5)
	SPEC (%)	38.7 (21.8 ; 57.8)	77.4 (58.9 ; 90.4)***	77.4 (58.9 ; 90.4)***	74.2 (55.4 ; 88.1)***	74.2 (55.4 ; 88.1)***
	PPV (%)	29.6 (13.8 ; 50.2)	56.3 (29.9 ; 80.2)	53.3 (26.6 ; 78.7)	46.7 (21.3 ; 73.4)	42.9 17.7 ; 71.1)
	NPV (%)	92.3 (64.0 ; 99.8)	100 (85.8 ; 100)	96.0 (79.6 ; 99.9)	92.0 (74.0 ; 99.0)	88.5 (69.8 ; 97.6)
	AUROC	0.62 (0.47 ; 0.78)	0.89*** (0.81 ; 0.96)	0.83*** (0.70 ; 0.96)	0.76 (0.60 ; 0.92)	0.70§ (0.52 ; 0.89)

_* = p < 0.05 with reference to EU, § = p < 0.05 with reference to APDI, *** = p < 0.001 with reference to EU, 20_CFI = CFI at the 20% offset, 22_APDI = APDI at the 22% offset

4.4 Discussion

The present study sought to determine the diagnostic value of AngioPLUS in differentiating vascular benign and malignant thyroid nodules. The diagnostic performance of sole approaches and qualitative and quantitative regional vascularity assessment approaches in combination with EU TIRADS in stratifying thyroid nodules was evaluated.

4.4.1 Sole TIRADS assessment in vascularized nodules

The present study initially assessed the diagnostic performance of EU TIRADS alone based on both the CAD and subjective approaches as was established in Chapter 3. EU TIRADS using subjective assessment was affirmed to be the optimal approach for the subsequent combination analyses due to higher sensitivity, specificity, and diagnostic efficacy overall (SEN: 83.3% vs 76.7%, SPEC: 50% vs 45.3%, and AUROC: 0.67 vs 0.61). The high sensitivity with lower specificity of the EU TIRADS can be attributed to its criteria of the presence of any single predictive feature to denote high malignancy risk/suspicion.

Few studies have evaluated the role of EU-TIRADS in nodules with equivocal cytology. The present study showed a sensitivity of 88.9% and a low specificity of 38.7% in stratifying nodules with equivocal cytology. Contrarily, a recent study demonstrated that EU TIRADS had lower sensitivity of 51.4% and 52.6% but higher specificity of 81.6% and 69.6% in FLUS/AUS nodules and SFN/HC nodules respectively [290]. Similarly, another study found EU TIRADS not very helpful in guiding the management of follicular or Hurthle cell neoplasms (SEN: 44.8%; SPEC: 78.9%) [291]. The differences between our findings and the other studies can be accounted for by the smaller sample size, the use of computer-assisted EU TIRADS stratification, lack of subgrouping of the equivocal nodules, and PTC as the predominant cancer in our study. The results of the present study can therefore be inferred as applicable to

PTCs and because of our lower sample size larger multi-centre studies with a variable distribution of cancers are necessary to validate and generalize these findings.

4.4.2 Qualitative vascularity assessment in combination with EU TIRADS

The present study demonstrated that qualitative assessment using the AngioPLUS Doppler technique alone and in combination with EU-TIRADS has higher diagnostic performance in the differentiation of benign and malignant nodules than conventional Doppler modes.

4.4.2.1 Qualitative Vascularity Assessment with conventional Doppler modes

This present study's findings of low sensitivity with high specificity using conventional colour and power Doppler techniques are consistent with those in some previous studies. Hong *et al.,*[160] found that prominent intranodular vascularity as a malignancy risk feature in CDU had a sensitivity of 31% and a specificity of 81%. Similarly, Wettasinghe *et al*., [292] reported a sensitivity of 48.5% and specificity of 79.1% using internal vascularity in contrast with a sensitivity of 87.9% and specificity of 58.7% using any single suspicious grey scale ultrasound feature. Contrarily, in one study the sensitivity of CDU was as high as 93.6% with a specificity of 86.7%; while in another study the sensitivity was 77% with a specificity of 48% [293, 294]. Using PDI to assess intranodular flow, Rosario *et al.,* [165] concluded that PDI had no additional diagnostic value in malignancy risk stratification. Grey scale alone had a sensitivity that was higher than that of PDI alone whereas the specificity was lower (SEN: 88.7% vs 15%; SPEC: 68.2% vs 96%); whilst both sensitivity and specificity were comparable to those of the combined grey scale + PDI approach (SEN: 88.7% vs 89.4%; SPEC: 68.2% vs 66.4%).

Although one study concluded that a central or predominantly central vascularity pattern in PDI was an independent risk factor for thyroid malignancy, other studies reported that intranodular vascularity was not a reliable malignancy predictor as it was also prominent in benign nodules [161, 164, 166]. The differences in the classification of vascularity into either central

vascularity or intranodular vascularity in various studies contribute to the differences in findings. While the regional classification of vascularity into central and peripheral areas likely provides a more distinct analysis, intranodular vascularity is more general and may refer to any presence of internal flow. Because of the differences in vascularity classifications, the subjective vascularity pattern analysis approaches also differ thereby resulting in variable findings.

4.4.2.2 Qualitative vascularity assessment with the AngioPLUS microvascularity Doppler modes

The diagnostic performance of AngioPLUS in combination with EU-TIRADS at a cut-off category of 4 for malignancy risk stratification resulted in an improved the overall diagnostic performance of EU TIRADS alone in the stratification of all nodules in the present study.

The diagnostic value of AngioPLUS in thyroid nodule differentiation lacks exploration. Although a similar microvascularization technique, SMI, has been evaluated in several studies, related studies involving combined assessment with EU TIRADS are lacking. However, a study based on ACR TIRADS reported that the combination of SMI to ACR TIRADS improved the diagnostic performance of ACR TIRADS alone (SEN: 65.12% to 93.75%, SPEC: 93.02% to 94.37% AUROC: 0.88 to 0.95,) [295]. The differences in the improved diagnostic performance outcomes between that study and ours may be explained by the difference in malignancy risk stratification at the selected cut-off categories. EU-category 4 is consistent with intermediate suspicion on a pattern-based approach, while that study used ACR category 4 which is consistent with moderate suspicion on a score-based approach. However, both our study and the previous study resulted in an improved overall diagnostic performance with the combination approaches of AngioPLUS + EU and SMI + ACR, respectively, than either TIRADS alone. Although there was a non-significant drop in sensitivity in our study, the improved overall diagnostic performance suggests that the inclusion of vascularity assessment

using AngioPLUS in the ultrasound assessment of thyroid nodules can improve diagnostic efficacy.

Another study reported that SMI improved the sensitivity of PDI (41.8% to 75.9%) and the combination of SMI with grey scale ultrasound features in a logistic regression model resulted in excellent diagnostic performance overall (AUROC: 0.92) [177]. Contrastingly, one study concluded that the combination of SMI with KSThR TIRADS did not significantly improve the diagnostic performance of grey scale assessment alone, while another study made the same conclusion based on the combination of any suspicious feature with SMI [179, 296]. Both studies varied from the present study in their scoring of the vascularity pattern using a 3-categories criterion for which intranodular vascularity was inclusive of both mild and extensive vascularity. This may explain the differences in the findings in addition to the different TIRADS criteria and that microvascularity assessment using SMI was based on the highly sensitive monochromatic mode.

4.4.2.3 Qualitative vascularity assessment in nodules with equivocal cytology

The additional diagnostic value of microvascularity imaging in thyroid nodules with equivocal cytology has not yet been examined even with techniques similar to AngioPLUS like SMI. In the present study, the combination of AngioPLUS-PDI with EU TIRADS category 4 improved the diagnostic performance over that of EU TIRADS category 5 alone, by maintaining the sensitivity while significantly increasing the specificity and overall diagnostic efficacy. From these results, we can posit that based on increased central or predominantly central vascularity detected with AngioPLUS and using an EU TIRADS cut-off criterion of intermediate suspicion, it reduces false-positive rates than sole EU TIRADS using the high suspicion category. The combination of the two ultrasound modes results in a diagnostically optimal

approach that significantly improves specificity while maintaining a high sensitivity which may potentially limit unnecessary biopsy rates in nodules with equivocal cytology.

A few studies have suggested that colour Doppler is beneficial in predicting malignancy in follicular neoplasms, and in one meta-analysis study, the pooled sensitivity and specificity of predominant internal flow were reported as 85% and 86% respectively [297, 298]. Our present study findings affirm this assertion suggesting that increased central vascularity detection using AngioPLUS is predictive of malignancy. This is attributed to the detection of microvasculature which cannot be achieved using conventional colour and power Doppler techniques. The high malignancy prediction with the increased detection of intranodular vascularity in nodules with equivocal cytology has been alluded to most cancers within these nodules being follicular thyroid carcinomas or their variants [180]. However, similar to a previous study that reported 79% PTCs and 20% FTCs in nodules with equivocal cytology, in the present study there were 7 PTCs (78%) and 2 FTCs (22%) in this group of nodules [4]. This is explained by most thyroid carcinomas usually being PTCs. However, the sequela of increased microvascularity because of the carcinogenesis process is likely anticipated more with FTCs since they have a hematogenous route of spread. Future larger studies with a greater distribution of FTCs and PTCs may help adequately establish the factors that influence nodule vascularity characteristics based on histology.

4.4.3 Quantitative vascularity assessment in combination with EU TIRADS

The present study demonstrated statistically significant differences between vascularity indices of sole colour and power Doppler modes and both modes coupled with AngioPLUS in the quantification of central and peripheral vascularity of benign and malignant nodules (*Figure 4:5 and Figure 4:6*). The higher quantification of vascularity with AngioPLUS, therefore, substantiates that there is increased sensitivity in microvascularity detection in thyroid nodules with this technique than conventional Doppler modes.

For the first time, in this study we employed a regional vascularity ratio analysis for thyroid nodule differentiation, using an RVI > 1 to denote predominantly peripheral vascularity. Our findings affirmed that benign nodules had predominant peripheral vascularity regardless of the offset used (*Table 4.4*). The diagnostic performance of the RVI method in analysing all nodules was substantial with CFI at the 20% offset, however; the diagnostic performance was generally comparable with other modes at the same offset except for PDI. Because the ratio is a relative measure of the peripheral VIs to that of central VIs, when there are marginal differences in either central or peripheral VIs between different offsets, the result is a minimal change in the ratio value for different offsets. The similar diagnostic performance outputs at different offsets can be attributed to this.

Thyroid nodule vascularity has been quantitatively evaluated in a few previous studies and shown variable results due to the differences in methodologies and cut-off criteria. Lyshchik *et al.*, [255] suggested that vascularity quantification was influenced by nodule size and achieved perfect specificity (100%) but lower sensitivity using normalized and weighted VIs (72.4% and 69%, respectively). In another study, the overall VI, central VI, and peripheral VI for vascularity densities yielded high sensitivities and low specificity at different high cut-off points (84.8% vs 40.9%, 83.5% vs 41.5%, and 78.5% vs 40.3%, respectively) [254]. Furthermore, that study reported that neither peripheral nor intranodular vascularity was predominant in benign or malignant nodules, although quantitatively benign nodules were more vascular than malignant nodules. While our present study used regional vascularity segmentation at various prescribed offsets to delineate peripheral and central regions of thyroid nodule and determine the best offset in ratio analysis, that study was based on 90% vs 10% central and peripheral regional segmentation, respectively.

Sultan *et al.*, [253] compared qualitative vascularity assessment and quantitative central vascular area and central flow volume index and reported a high sensitivity with comparable specificity

using quantitative central vascular area than qualitative vascularity assessment (SEN: 90% vs 67.5%, SPEC: 88.1% vs 88%). However, like the other mentioned studies the combined performance with grey scale ultrasound was not assessed. A previous study concluded that quantitative vascularity assessment using power Doppler ultrasound was poor at thyroid malignancy prediction based on the low diagnostic performance of sole VI assessment (SEN: 53.7% and SPEC: 58.8%) [171]. In that study, although the combination of VI with grey scale ultrasound assessment improved the overall diagnostic efficacy, the comparable sensitivity and significantly reduced specificity suggested no improvement of the overall performance of grey scale ultrasound alone. (SEN: 95.8% vs 89%, $p > 0.05$; SPEC: 28.2% vs 74.3% and AUROC: 0.82 vs 0.70, $p < 0.001$).

In a previous study using regional vascularity segmentation, our group established that the combination of VIs (overall, central, and peripheral) at the 22% offset with grey scale assessment improved the specificity but resulted in a reduced sensitivity from that of grey scale assessment alone (SEN: 96.3% to 66.7% and SPEC: 46.4% to 83.3%). Because the present study employed the same regional vascularization principle although ratio analysis was used; the outcomes were similar for the combined approach resulting in a reduced sensitivity from 83.3% to 53.3% but increased specificity from 50% to 81.3% when compared to grey scale ultrasound alone. Ultimately, the ratio analysis approach of RVI > 1 to denote predominant peripheral vascularity may accurately stratify benign nodules but it is not ideal for ruling in disease.

4.4.3.1 Quantitative vascularity in nodules with equivocal cytology

Quantitative vascularity assessment of thyroid nodules with equivocal cytology is lacking. In this first-time assessment, the ratio analysis approach seems to perform best in this subgroup of nodules at an offset of 22% and 25% using PDI coupled with AngioPLUS. The sole APDI

approach resulted in optimal sensitivity and specificity, while the combined approach with EU TIRADS resulted in a reduced sensitivity but sustained specificity. Although there is potential to reduce unnecessary biopsy rates with this approach because of improved specificity, the large sacrifice of the sensitivity would be a drawback in clinical application considerations. This is because reduced sensitivity of a diagnostic method would increase the false-negative rate which may delay treatment of thyroid cancer patients. The qualitative approach in combination with EU-TIRADS is, therefore, more suitable for best stratifying thyroid nodules with equivocal cytology.

4.5 Limitations

This study had several limitations. The sample size was small with a high prevalence of PTC among the malignant nodules thereby limiting the generalisability of the findings to other cancers. Furthermore, the sample size of the subgroup of nodules with equivocal cytology was small to conduct analyses of the equivocal cytology categories 3 and 4 separately. The separate evaluation could help identify which cytology category the microvascularity analysis would be of the most additional clinical value. Due to the selection of patients with FNAC and/or histopathology results, we cannot exclude selection bias. Because of the currently limited evaluation of the diagnostic performance of AngioPLUS along with several TIRADS and multiple raters, future larger prospective studies are warranted to validate our findings.

4.6 Conclusions

This study evaluated the diagnostic value of microvascularity assessment with AngioPLUS in combination with EU TIRADS based on qualitative and quantitative vascularity assessment. We established that although quantitative regional vascularity based on RVI could adequately discriminate benign nodules, it has little value in ruling in malignancy. Therefore, the qualitative assessment of vascularity with AngioPLUS coupled with power Doppler ultrasound in combination with EU TIRADS has high diagnostic efficacy in differentiating benign and malignant nodules. This combined approach has more potential for improving the diagnostic accuracy in diagnosing thyroid nodules with equivocal cytology as it results in improved specificity without compromising the sensitivity.

The findings in this study have the potential for clinical application. The combined approach of EU TIRADS and AngioPLUS with power Doppler has potential value as a follow-up ultrasound method for thyroid nodules with equivocal cytology. Because the approach seems sensitive enough to detect any cancers and also has high specificity that limits unnecessary repeated biopsy, it may result in a more conservative diagnosis of thyroid nodules.

Chapter 5

Study Three: The Diagnostic Value of Shear wave Elastography in Combined Assessment with EU TIRADS and Multi-modal Ultrasound Assessment with AngioPLUS for Thyroid Nodule Differentiation

5.1 Introduction

Elasticity imaging has been suggested as a useful complementary imaging modality that can improve the specificity and overall accuracy of grey scale ultrasound assessment in thyroid nodule diagnosis in addition to FNAC assessment [182, 183]. Quantitative approaches such as shear wave elastography that result in absolute tissue stiffness values rather than relative values or ratios are purported to be more objective and less user-dependent than strain elastography.

FNAC is regarded as the pre-surgery reference standard for the diagnosis of thyroid nodules. While its sensitivity and specificity can both range up to over 90%, it has challenges of non-diagnostic results in about 10% of the samples, and equivocal results in up to 30% of cases [107, 147]. Grey scale ultrasound remains the first-line pre-operation diagnostic imaging method for thyroid cancers. However, due to the overlap of some ultrasound features in benign and malignant nodules, no sole feature is highly predictive of malignancy [4]. Different TIRADS have evolved to aid malignancy risk estimation and FNAC selection of thyroid nodules based on multiple grey scale ultrasound features and nodule size. Nonetheless, the diverse malignancy risk stratification criteria amongst different TIRADS result in varying diagnostic accuracy outcomes, thereby causing the lack of a universal standard for clinical use. The evaluation of thyroid nodule stiffness as indicative of malignancy is purported to result in higher diagnostic accuracy when compared to different TIRADS [288, 299].

Although SWE has gained popularity in recent years, its utility for thyroid cancer detection has been hampered by the lack of standardized diagnostic criteria. There is a lack of consensus

regarding the best SWE measurement parameters, corresponding cut-off points and the SWE measurement techniques for the best diagnostic efficiency. Recent meta-analysis data showed different optimal cut-off ranges of between 27.7 to 85.2kPa for the mean SWE index which corresponded with sensitivity and specificity ranges of 53% to 95% and 70% to 99%, respectively [191]. In other studies, the maximum SWE index was reported as having the best diagnostic efficiency while the minimum and SD indices were seldom reported as achieving the best diagnostic performance [26, 31, 300, 301]. There is also conflicting evidence regarding the benefit of combining SWE with grey scale ultrasound features or TIRADS in the diagnosis of thyroid cancer. Some studies reported an improved diagnostic performance with combined assessment, while others suggested that grey scale ultrasound feature assessment alone sufficed [302-305]. Although SWE has been suggested to have diagnostic value in aiding the diagnosis of thyroid nodules with equivocal cytology [32, 181], the limited current studies exhibit similar challenges.

Multi-modal ultrasound assessment has been proposed as a solution to improving the overall diagnostic efficacy in differentiating thyroid nodules, more so nodules with equivocal cytology [146]. Molecular tests are most accurate in diagnosing nodules with equivocal cytology; however, they are very expensive and not easily accessible [80]. Different approaches of vascularity and elasticity ultrasound assessment in addition to grey scale ultrasound assessment, prevail in the few current multi-modal thyroid ultrasound studies. Typical multi-modal assessment often involves either 2D/3D/4D-ultrasound, colour Doppler ultrasound assessment, superb microvascular imaging, or CEUS combined with strain elastography assessment and variable grey scale ultrasound assessments [178, 200, 201]. The diversity of methodological approaches in different studies has resulted in conflicting opinions regarding the diagnostic value of multi-modal ultrasound imaging in thyroid nodule diagnosis. Furthermore, the diagnostic value of multimodal approaches involving SWE, AngioPLUS microvascularity assessment in addition

to specific TIRADS is still uncertain and needs further exploration for evidence-based clinical practice.

The present study sought to evaluate the diagnostic efficiency of SWE in combination with EU TIRADS (the best TIRADS based on the results of Chapter 3) in thyroid nodule diagnosis. Taking into consideration of the potential influencing factors affecting the diagnostic performance of SWE, all nodules were evaluated followed by subcategory analyses of nodule size, equivocal cytology, and vascularity status. The scan plane and SWE parameters that best discriminated benign and malignant nodules with statistically significant differences were used for the diagnostic performance assessments. Furthermore, a multi-modal approach involving EU TIRADS, AngioPLUS microvascularity assessment and SWE was evaluated to determine its diagnostic value in differentiating thyroid nodules with equivocal cytology. The findings of the present study will broaden the understanding of the clinical value of SWE–multi-modal approaches for thyroid nodule diagnosis. This in turn could potentially facilitate efficient patient management in the clinical utility of SWE.

5.2 Materials and methods

5.2.1 Study type

This was a prospective analytical observational study that received ethical approval from the Institutional Human Research Ethics sub-committee before it was conducted. Based on consecutive case analysis approach and non-probability sampling, cross-sectional cohorts of patients with thyroid nodules and/or suspicion of thyroid cancer were purposively recruited at the Prince of Wales Hospital Department of Surgery and its affiliates from 15 May 2019 to 31 August 2021. Informed consent was sought from the patients before the data collection procedures.

5.2.2 Data collection procedures

5.2.2.1 Inclusion and exclusion criteria

The inclusion criteria in this study were all consenting adult patients (≥18 years old) who had thyroid nodular disease or suspicion of thyroid cancer and were scheduled for FNAC and/or subsequent surgery. Nodules that were between 5mm and 50mm were included in the study. The lower limit for the nodule size was guided by the computer-assisted subjective ultrasound assessment approach that was used whose limit was ≥ 5mm. This is consistent with the size criteria for FNAC recommendation in addition to the clinical or sonographic risks when using most TIRADS. The upper limit was determined as the largest size that the transducer footprint can cover and can also be completely encompassed by the SWE colour overlay. For patients with multiple thyroid nodules, either the nodule with the most suspicious sonographic features (hypoechoic, microcalcifications, irregular margins, tall-than-wide, etc.), if there were no obvious suspicious features then the largest nodules, or the one/s for which biopsy and/or surgery were recommended, were included in the study. The exclusion criteria were patients < 18 years old, patients who had a thyroidectomy, did not have a conclusive diagnosis as

determined from either cytology results, histopathology results or both, and multinodular goitres without clearly isolated sole nodules. Patients with only completely cystic nodules or nodules that were too large for the footprint of the transducer and cannot be completely visualized in the image field of view were excluded from the study as they would affect the elastography output [183].

5.2.2.2 Ultrasound imaging procedures

A sole investigator (N.C.) conducted the thyroid ultrasound imaging of all patients using the same Aixplorer ultrasound machine (Supersonic Imagine, Aix-en-Provence, France) equipped with a 7-10 MHz linear transducer, AngioPLUS and SWE. The ultrasound machine settings were standardized during a pilot study and the same ultrasound scanning preset for thyroid study was maintained throughout the study to ensure consistency.

The investigator followed standard ultrasound scanning protocols to conduct the thyroid scans. Each patient lay in the supine position with minimal extension of the neck, and coupling gel was applied. With the face turned away from the side of interest, and the transducer placed on the exposed side of the neck. A minimum of 3 transverse and 3 longitudinal images of each target thyroid nodule in grey scale, colour Doppler, power Doppler, colour Doppler and power Doppler coupled with AngioPLUS, and SWE modes were acquired.

For the SWE mode, a generous layer of coupling gel was applied on the patient's neck to minimize transducer compression on the neck and the static SWE images were acquired on arrested inspiration. The pre-set quantitative SWE measurement scale on the ultrasound machine ranged from 0 to 180 kPa. The SWE sampling box was adjusted to cover the whole thyroid nodule and the trace mode was used to manually outline the ROI. Images with intranodular cystic areas, calcified areas and areas void of SWE colour were avoided where possible. The inbuilt quantification tool, Q-box, automatically computed the minimum,

maximum, mean and standard deviation (SD) elasticity indices in Young's modulus (kilopascals- kPa) for each of the 3 images in the transverse and longitudinal planes (*Figure 5:1*). In the present study, the average values for each SWE index in kPa from the 3 images for each scan plane were compared against the final histopathology results to determine diagnostic accuracy. The index and scan plane that demonstrated statistically significant differences between benign and malignant thyroid nodules were used for the diagnostic performance assessment.

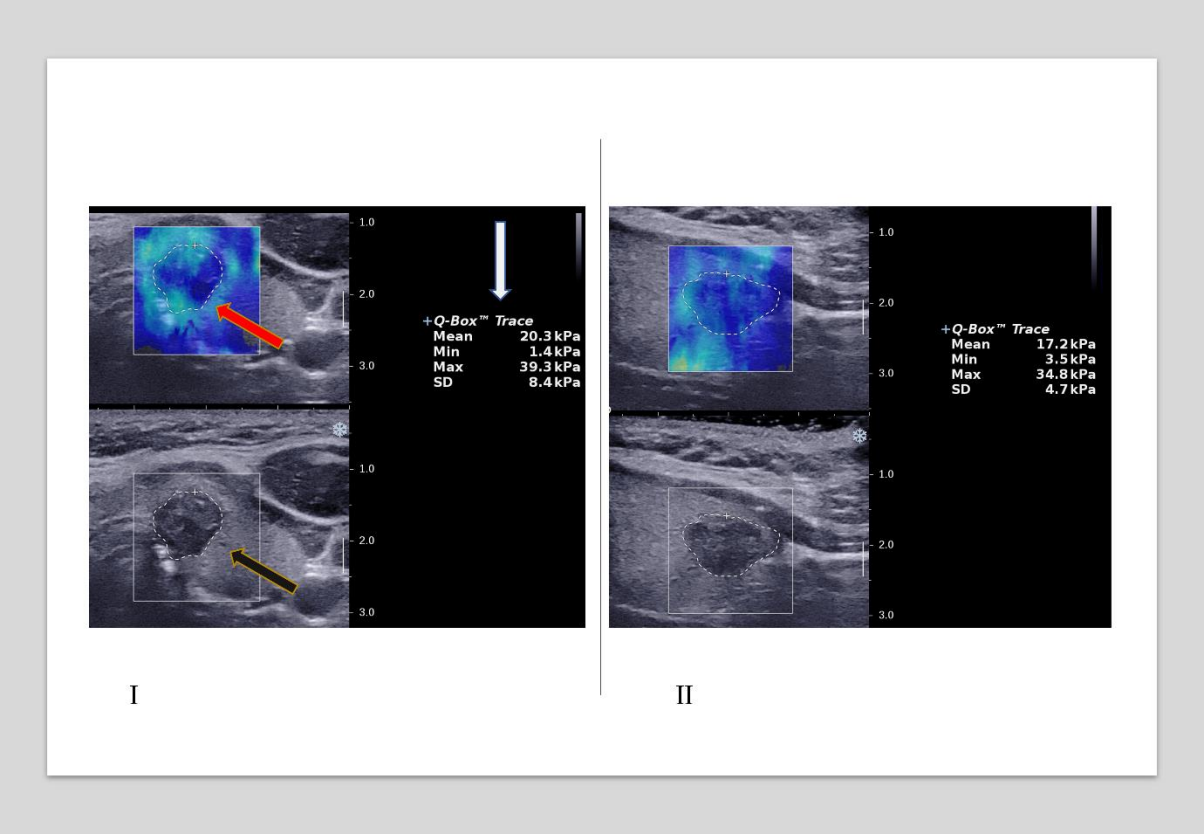


Figure 5:1: Image representation of the ROI selection for the quantification of stiffness in SWE. I) An illustration of the Q-box placement and the nodule trace outline (red arrow), the output of the quantification of thyroid nodule elasticity in different SWE indices (white arrow) and the grey scale overlay image of the same nodule (black arrow) in the transverse plane. II) The ROI trace representation and SWE quantification of the same nodule in the longitudinal plane.

5.2.2.3 Image analysis procedures

A stepwise approach was used to analyse the images based on each ultrasound modality used. The initial step involved analysing all thyroid nodules using grey scale ultrasound and SWE. Images were then distributed into relevant subcategories based on nodule size, equivocal cytology, and vascularity status. Figure 5:2 illustrates the analysis sequence and criteria. Nodules that had FNAC classifications 3 and 4 (atypia of undetermined significance or follicular lesion of undetermined significance – AUS/FLUS; and suspicion of follicular neoplasm or follicular neoplasm – SFN/FN, respectively) were considered equivocal. The sole grey scale ultrasound assessment was based on EU TIRADS at a pre-determined cut-off point of category 5 as previously established in Chapter 3 (*Section 3.4.2.5*) since the same computer-assisted subjective approach was used in this study. The AngioPLUS mode coupled with power Doppler (APDI) was used for the vascularity assessment, and predominantly central and exclusively central vascularity were considered suggestive of malignancy.

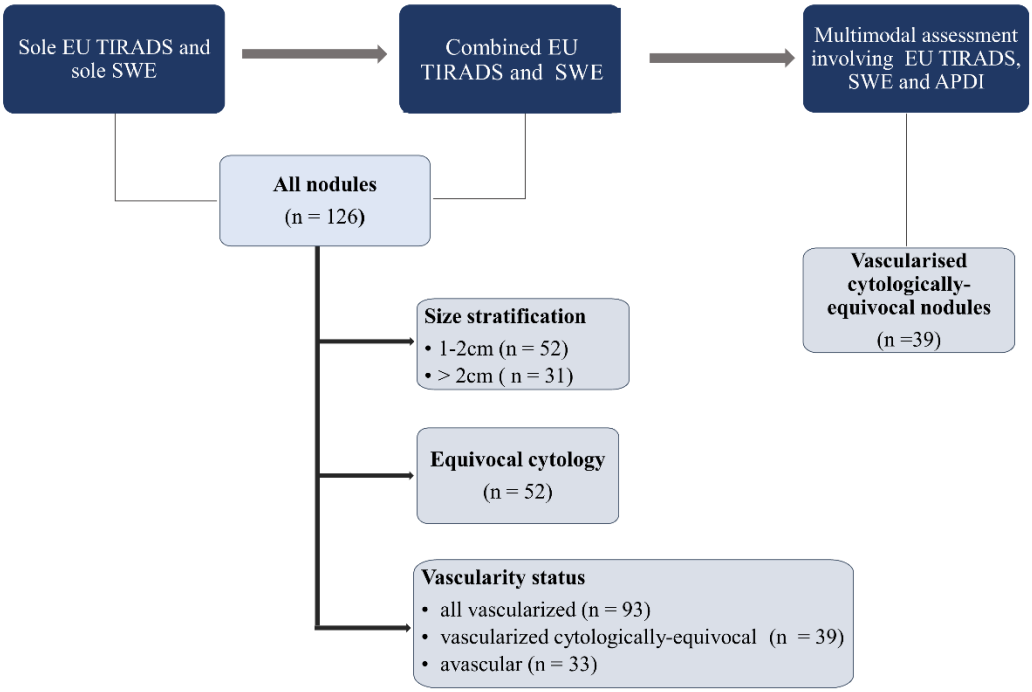


Figure 5:2: An illustration of the steps that were involved in the SWE analyses of the thyroid nodule images.

5.2.3 Data analysis and statistical analysis

Continuous data was classified as means +/- SD whereas categorical and/or nominal data was expressed as frequencies and percentages. The Chi-Square test was used to compare differences in nodule classification data. The Shapiro Wilk test was used to check the normality of the data. The Wilcoxon ranks test was used for the paired comparison of the SWE measurements between transverse and longitudinal planes while the Mann-Whitney U test was used to compare the SWE measurements between benign and malignant nodules.

The SEN, SPEC, PPV, NPV and DOR were calculated with reference to final cytology or histopathology results. The ROC curves were generated to obtain the AUROC and the optimal cut-off points for the SWE measurements were determined to be the maximum value of the sum of the specificity and the sensitivity of the AUROC. For the different subcategories of the thyroid nodules, the diagnostic performance measures were determined as follows:

i) sole EU TIRADS and the average of each of the mean, maximum, minimum SD SWE indices

ii) combination of EU TIRADS + each of the SWE indices at the determined cut-off values

iii) sole EU TIRADS and each of the statistically significant SWE indices at the determined cut-off values for the different subgroups of the nodules

iv) combination of EU TIRADS + SWE indices for the different subgroups of the nodules

v) multi-modal assessment of EU TIRADS, SWE and APDI

The optimal cut-off point for sole EU TIRADS was pre-determined to be category 5 as was previously established. For the combined assessment of TIRADS and SWE, a nodule was

suspected of malignancy if it met the cut-off criteria of the SWE and an EU TIRADS cut-off of ≥ 4. For the multi-modal assessment of EU TIRADS, SWE and APDI, a nodule was suggestive of malignancy if it met the EU TIRADS cut-off of ≥ 4, and either demonstrated predominant/exclusive central vascularity on APDI or met the SWE cut-off criteria. The McNemar and Cochran Q's tests were used for the comparative analysis of sensitivity and specificity whereas the z-test was used to compare the different AUROCs. The tests were two-sided and $p < 0.05$ denoted statistical significance. The SPSS statistical software (version 26.0, SPSS Inc., Chicago, IL, USA) was used for the analyses.

5.3 Results

5.3.1 Demographics and nodule classification data

In this study, 146 patients were recruited and scanned. Twenty-five patients were excluded as they did not comply with the study inclusion criteria. Finally, a total of 126 thyroid nodules (81 benign and 45 malignant) from 121 patients (100 females and 21 males) were included in the study. Figure 5:3 shows the patient and thyroid nodule selection steps and exclusion reasons. The mean age of the patients included in the study was 53.8 ± 12.8 (range: 27 to 75) years. The mean age of male patients (62.1 ± 8.8, range: 44 to 73) was significantly higher than the mean age of female patients (52.1 ± 12, range: 27 to 75), $p < 0.001$. Table 5.1 shows the demographic data results of the patients. The mean nodule size was not significantly different between benign nodules (1.6 ± 0.8 cm, range: 0.5 to 3.6 cm) and malignant nodules (1.3 ± 0.8 cm, range: 0.5 to 3.7 cm), $p > 0.05$. The predominant nodule size category was 1 to 2cm for benign nodules (65.4%) and < 1cm for malignant nodules (46.5%). The classification of the nodules based on the cytology category revealed that 52 nodules (45.2%) had equivocal cytology of which 37 (71.2%) had benign histopathology results. The false-negative rate based on the FNAC was 4% (2/45).

The common histopathology diagnosis of the malignant nodules was papillary thyroid carcinoma (PTC, n = 39), while the remaining nodules were classified as non-invasive follicular thyroid neoplasm with papillary like nuclear features (NIFTP, n = 2), follicular variant of PTC (FvPTC, n = 1), follicular thyroid carcinoma (FTC, n = 1), widely invasive FTC (n = 1) and minimally invasive FTC (n = 1). Most of the benign nodules had unspecified benign histopathology (n = 27) followed by nodular goitres (n = 16) while the rest were nodular hyperplasia (n = 6), Hurthle cell adenoma (n = 5), follicular adenoma (n = 4), Graves' disease (n = 4), colloid nodules (n = 3), Hashimoto's thyroiditis (n = 2) and lymphocytic thyroiditis (n = 2). The remaining 12 nodules had a benign diagnosis based on FNAC only, 8 of which did

not undergo surgery while the other 4 underwent a radiation frequency ablation (RFA) procedure.

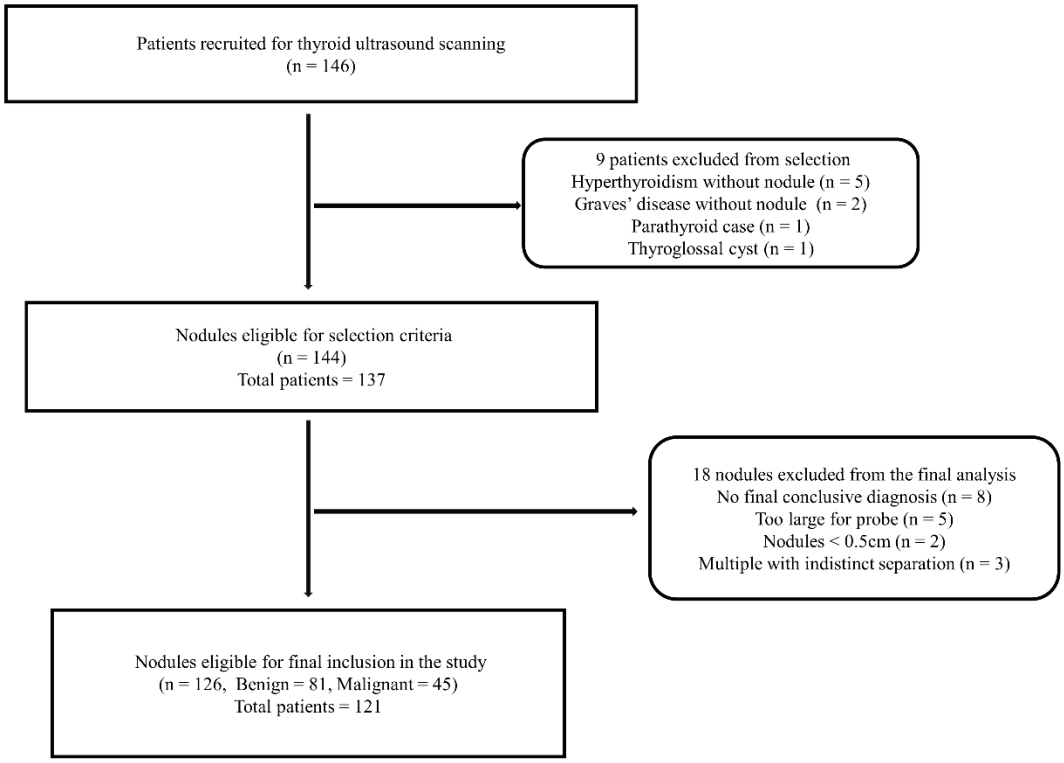


Figure 5:3: The image selection process of the ultrasound images of the thyroid nodules for the SWE study

Table 5.1:Demographic data and the distribution of nodules into different classifications

		Mean/ Frequency by Diagnosis		
Characteristic	**Overall mean/ frequency**	**B**	**M**	**p-value**
Gender				
Male	21	14 (66.7 %)	7 (33.3%)	> 0.05
Female	100	67 (67%)	33 (33%)	
Mean Age				
Overall	53.8 ± 12	53.8 ± 12.1	54.0 ± 12	> 0.05
Male	62.1 ± 8.8			< 0.001
Female	52.1 ± 12			
Nodule size				
Total nodules	126	81 (64.3%)	45 (35.7%)	< 0.01
Overall mean(cm)	1.5 ± 0.8	1.6 ± 0.8	1.3 ± 0.8	0.62
< 1cm	43	23(53.5%)	20 (46.5%)	0.10
1-2cm	52	34 (65.4%)	18 (34.6%)	
>2cm	31	24 (29.6%)	7 (22.6%)	
FNAC				
Not done	11	11(100%)	0 (0%)	< 0.001
Non-diagnostic	6	5 (83.3%)	1(16.7%	
Benign	30	28 (93.3%)	2 (6.7%)	
Equivocal	52	37 (71.2%)	15 (28.9%)	
Malignant/SOM	27	0 (0%)	27 (100%)	
EU TIRADS				
1	0	0 (0%)	0 (0%)	< 0.001
2	27	24 (88.9%)	3 (11.1%)	
3	0	0 (0%)	0 (0%)	
4	22	18 (81.8%)	4 (18.2%)	
5	77	39 (50.7%)	38 (49.4%)	

B = benign; M = malignant; SOM = suspicion of malignancy, FNAC = fine needle aspiration cytology

5.3.2 Analysis of the different SWE indices in thyroid nodule differentiation

5.3.2.1 Analysis based on the different scan planes

The medians of the different SWE indices were compared between transverse plane and longitudinal plane measurements for all nodules. The Wilcoxon signed rank test revealed that SWE measurements were significantly lower with the transverse plane for the mean and minimum indices than with the longitudinal plane (mean SWE index: median = 15.1 kPa vs 17.9 kPa, $z = -4.61$, $p < 0.001$; minimum SWE index: median = 0.2 kPa vs 1.4 kPa, $z = -6.06$, $p < 0.001$, respectively). The medians for the maximum and the SD SWE indices were not significantly different between the transverse and longitudinal scan measurements (maximum SWE index: median = 43.6 kPa vs 42.3 kPa, $z = -0.58$, $p = 0.56$; SD SWE index: median = 7.2 kPa vs 7.4 kPa, $z = -0.27$, $p = 0.79$, respectively).

5.3.2.2 Analysis of nodule stratification based on different sub-categories

The differences of the medians of the SWE measurement indices between benign and malignant nodules were evaluated with the Mann Whitney- U test based on the different subcategories of the nodules. Table 5.2 shows the p values of the statistical analyses of different SWE measurement indices and nodule subcategories. The transverse and longitudinal mean measurements, the longitudinal minimum, and the transverse SD measurement indices were statistically significant between benign and malignant nodules for all nodules, nodules of sizes between 1 to 2cm and all vascular nodules ($p < 0.05$). The longitudinal minimum measurement index was statistically significant for the nodules with equivocal cytology, vascular nodules with equivocal cytology and avascular nodules ($p < 0.05$). The transverse SD measurement index was statistically different for nodules that were greater than 2cm ($p < 0.05$). All SWE measurement indices did not differ significantly between malignant and benign for the nodules that were less than 1cm ($p > 0.05$).

Table 5.2: Statistical significance assessment for the differences in SWE indices between benign and malignant thyroid nodules based on the scan planes

Nodule category	**p-values of SWE indices in kPa**							
	T_{Mean}	L_{Mean}	T_{Min}	L_{Min}	T_{Max}	L_{Max}	T_{SD}	L_{SD}
All N = 126 (B = 81, M = 45)	0.005**	0.007**	0.100	0.003**	0.061	0.253	0.012*	0.255
Equivocal N = 52 (B = 37, M = 15)	0.473	0.214	0.353	0.015*	0.313	0.391	0.138	0.525
< 1cm N = 43 (B = 23, M = 20)	0.189	0.141	0.128	0.077	0.368	0.480	0.219	0.733
1-2cm N = 52 (B = 34, M = 18)	0.017*	0.010*	0.865	0.195	0.034*	0.102	0.009**	0.108
>2cm N = 31 (B = 24, M = 7)	0.104	0.661	0.216	0.835	0.061	0.417	0.029*	0.085
Vascular (all) N = 93 (B = 60, M = 33)	0.023*	0.026*	0.269	0.019*	0.132	0.324	0.041*	0.368
Vascular (eqv) N = 39 (B = 30, M = 9)	0.756	0.402	0.909	0.025*	0.806	0.781	0.384	0.831
Avascular N = 33 (B = 17, M = 16)	0.146	0.127	0.179	0.031*	0.581	0.763	0.423	0.929

SWE = shear wave elastography, B = benign, M = malignant, T_{Mean} = transverse Mean, LMean = longitudinal Mean, T_{Min} = transverse Minimum, L_{Min} =longitudinal Minimum, T_{Max} = transverse Maximum, LMax = longitudinal Maximum, T_{SD} = transverse Standard deviation, L_{SD} = longitudinal Standard deviation, EU = European, eqv = equivocal
*$p < 0.05$ = *, $p < 0.01$ = ***

5.3.3 Diagnostic performance assessment of the SWE indices with EU TIRADS based on the different subcategories of the nodules

This present study evaluated the diagnostic performances of sole grey scale ultrasound assessment with EU TIRADS, sole SWE measurement indices, and combined EU TIRADS and SWE for the different subcategories. The optimal cut-off points of SWE indices that showed statistically significant differences between malignant and benign nodules (*Table 5.3*) were determined and used in the diagnostic performance assessment.

5.3.3.1 Assessment of all nodules

The diagnostic performance of sole EU TIRADS showed the highest sensitivity (84.4%), lowest specificity (51.9%) and the highest acceptable diagnostic efficacy (AUROC: 0.69) overall in diagnosing all nodules. The diagnostic performance results are shown in Table 5.3. At the optimal cut-off, the transverse mean (T_{Mean}) SWE index resulted in diagnostic performance outcomes that were comparable to those of the longitudinal mean (L_{Mean}) SWE index (SEN: 51.1% vs 42.2%, and SPEC: 77.8% vs 88.9%, $p > 0.05$). The diagnostic performance outcomes of the transverse SD (T_{SD}) and the longitudinal minimum (L_{Min}) indices (SEN: 51.1% and 53.3%, and SPEC: 76.5% and 76.5%, respectively) were comparable to those of both mean indices ($p > 0.05$). All the sole SWE indices had significantly lower sensitivity but higher specificity than the sole EU TIRADS ($p < 0.001$).

The combined EU TIRADS and SWE diagnostic performance assessment still resulted in significantly lower sensitivities and higher specificities than EU TIRADS alone ($p < 0.001$). The overall diagnostic efficacy of sole EU TIRADS remained the highest (AUROC: 0.69) but without a statistically significant difference from any of the combined approaches ($p > 0.05$).

5.3.3.2 Assessment of nodules based on nodule size

The diagnostic performance of EU TIRADS and SWE measurements was evaluated based on nodules that were 1 to 2cm in size and those greater than 2cm.

The diagnostic performance outcomes of nodules that were 1 to 2cm in size for EU TIRADS alone showed the highest sensitivity (88.9%) and the lowest specificity (55.9%). The T_{Mean} SWE index at the optimal cut-off point of 25.6 kPa resulted in the lowest sensitivity (50%) and highest specificity (94.1%). In contrast, the L_{Mean} SWE index with the optimal cut-off point of 23.4 kPa yielded a sensitivity of 66.7% and a specificity of 79.4%. Out of all the SWE indices, the T_{SD} at the optimal cut-off of 8.7 kPa had the highest sensitivity (77.8%) and lowest specificity (64.7%) which were not significantly different from those of EU TIRADS alone ($p > 0.05$). Combined EU TIRADS and SWE diagnostic performance assessment revealed that EU TIRAD + T_{SD} had the highest sensitivity of 72.2% which was lower but not statistically significantly different from EU TIRADS alone (88.9%, $p > 0.05$) and a specificity of 76.5% which was significantly higher than that of EU TIRADS alone (55.9%, $p < 0.01$). The overall diagnostic efficacy of EU TIRADS + T_{SD} was the highest although it was comparable to that of EU TIRADS alone (AUROC: 0.74 vs 0.72, $p > 0.05$).

For the subcategory of nodules greater than 2cm, the EU TIRADS maintained a high sensitivity but a lower specificity. The sensitivity was not significantly different from that of sole SWE T_{SD} at the optimal cut-off of 10.7 kPa and of the combined EU + T_{SD} approach (85.7% vs 71.4% and 71.4%, $p > 0.05$) while the specificity was significantly different (62.5% vs 83.3% and 95.8%, $p < 0.01$). The combined approach resulted in the highest diagnostic efficacy and comparably high predictive values (AUROC: 0.84, PPV: 83.3%, and NPV: 92%).

5.3.3.3 Assessment of nodules with equivocal cytology

The diagnostic performance assessment of the 52 nodules with equivocal cytology revealed that the EU TIRADS alone resulted in significantly higher sensitivity but lower specificity than the sole SWE L_{Min} at the optimal cut-off of 6.1kPa and the combined approach (SEN: 80% vs 60% and 60%, $p < 0.05$; SPEC: 37.8% vs 78.4% and 83.4%, $p < 0.001$). The overall diagnostic efficacy of the combined approach was significantly higher than that for EU TIRADS alone (AUROC: 0.72 vs 0.58, $p < 0.05$).

Table 5.3: Diagnostic performance assessment of sole and combined EU TIRADS and SWE indices based on all nodules, equivocal cytology, and size stratification

Nodule category	**Diagnostic Test**	**Optimal cut-off**	**SEN (%)**	**SPEC (%)**	**PPV (%)**	**NPV (%)**	**AUROC**
All	EU	5	84.4	51.9	49.4	85.7	0.69
	T_{Mean} (kPa)	19.3	51.1***	77.8***	56.1	74.1	0.65
	L_{Mean} (kPa)	28.2	42.2***	88.9***	67.9	73.5	0.65
	T_{SD} (kPa)	10.5	51.1***	76.5***	54.8	73.8	0.64
	L_{Min} (kPa)	4.7	53.3***	76.5***	55.8	74.7	0.66
	EU + T_{Mean}		48.9***	82.7***	61.1	74.4	0.66
	EU + L_{Mean}		40.0***	92.6***	75.0	73.5	0.66
	EU + T_{SD}		51.1***	84.0***	63.9	75.6	0.68
	EU + L_{Min}		51.1***	77.8***	56.1	74.1	0.64
Size 1-2cm	EU	5	88.9	55.9	51.6	90.5	0.72
	T_{Mean} (kPa)	25.6	50.0***	94.1***	81.8	78.0	0.70
	T_{Max} (kPa)	50.2	61.1**	73.5**	55.0	78.1	0.68
	T_{SD} (kPa)	8.7	77.8	64.7	53.9	84.6	0.72
	L_{Mean} (kPa)	23.4	66.7*	79.4**	75.0	83.3	0.72
	EU + T_{Mean}		44.4***	94.1***	80.0	76.2	0.69
	EU + T_{Max}		55.6***	82.4***	62.5	77.8	0.69
	EU + T_{SD}		72.2	76.5**	61.9	83.9	0.74
	EU + L_{Mean}		61.1**	85.3***	68.8	80.6	0.73
Size > 2cm	EU	5	85.7	62.5	40.0	93.8	0.73
	T_{SD} (kPa)	10.7	71.4	83.3**	55.6	90.9	0.77
	EU + T_{SD}		71.4	95.8**	83.3	92.0	0.84
Equivocal	EU	5	80.0	37.8	34.3	82.4	0.58
	L_{Min} (kPa)	6.1	60.0*	78.4***	52.9	82.9	0.64
	EU+ L_{Min}		60.0*	83.4***	60.0	83.8	0.72*

EU = European TIRADS, T_{Mean} = transverse Mean, L_{Mean} = longitudinal Mean, L_{Min} =longitudinal Minimum, T_{Max} = transverse Maximum, T_{SD} = transverse Standard deviation, kPa = kiloPascals
* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ relative to EU TIRADS

5.3.3.4 Assessment of thyroid nodules based on the vascularity status

The diagnostic performance of EU TIRADS and SWE measurements was evaluated based on nodule vascularity. Table 5.4 shows these results. For the 93 vascularized nodules, the EU TIRADS had the best overall diagnostic performance with the AUROC of 0.69 and achieved the highest sensitivity (86.7%) but lowest specificity (50.8%). The diagnostic performance outcomes of the different sole SWE indices were all comparable to each other. The combined approaches of EU TIRADS with different SWE indices had diagnostic performance outcomes which were comparable to those of the sole SWE indices with higher specificities but lower sensitivities than EU TIRADS ($p < 0.001$). The sensitivity of EU TIRADS was higher but not significantly different from that of L_{Min} and the combined approach in discriminating cytologically equivocal vascular nodules (SEN: 88.9% vs 66.7% and 66.7%, $p > 0.05$). Conversely, the specificities of L_{Min} and the combined approach were significantly higher with an overall higher but non-significant diagnostic efficacy than EU TIRADS (SPEC: 80% and 80% vs 36.7%, $p < 0.001$; AUROC: 0.73 and 0.73 vs 0.62, $p > 0.05$).

For discriminating the avascular nodules, the sensitivity of EU TIRADS was not significantly different from L_{Min} at the optimal cut-off of 0.9 kPa and the combined approach, but the specificity was significantly lower than L_{Min} and the combined approach (SEN: 81.3% vs 75% and 68.8%, $p > 0.05$; SPEC: 52.9% vs 76.5% and 76.5%, $p < 0.01$). The combined approach had the highest diagnostic efficacy but a slightly lower sensitivity overall but without statistically significant differences from the sole methods.

Table 5.4: Diagnostic performance assessment of sole and combined EU TIRADS and SWE indices based on the nodule vascularity status

Nodule classification	**Diagnostic Test**	**Cut-off**	**SEN (%)**	**SPEC (%)**	**PPV (%)**	**NPV (%)**	**AUROC**
Vascular	EU	5	86.7	50.8	45.6	88.9	0.69
	T_{Mean} (kPa)	19.3	50.0**	81.0**	55.6	77.3	0.66
	T_{SD} (kPa)	11.8	40.0***	87.3***	60.0	75.3	0.64
	L_{Mean} (kPa)	23.5	46.7***	84.1**	58.3	76.8	0.65
	L_{Min} (kPa)	6.1	46.7***	84.1**	58.3	76.8	0.65
	EU + T_{Mean}		46.7***	85.7***	60.9	77.1	0.66
	EU + T_{SD}		40.0***	92.1***	70.6	76.3	0.66
	EU + L_{Mean}		43.3***	87.3***	61.9	76.4	0.65
	EU + L_{Min}		43.3***	85.7***	59.1	76.1	0.65
Vascular	EU	5	88.9	36.7	29.6	91.7	0.62
(Equivocal)	L_{Min} (kPa)	6.1	66.7	80.0***	50.0	88.9	0.73
	EU + L_{Min}		66.7	80.0***	50.0	88.9	0.73
Avascular	EU	5	81.3	52.9	61.9	75.0	0.68
	L_{Min} (kPa)	0.9	75.0	76.5**	75.0	76.5	0.72
	EU + L_{Min}		68.8	76.5**	73.3	72.2	0.73

EU = European TIRADS, T_{Mean} = transverse Mean, T_{SD} = transverse Standard deviation, L_{Mean} = longitudinal Mean, L_{Min} = longitudinal Minimum, kPa = kiloPascals
* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$ relative to EU TIRADS

5.3.4 Multi-modal diagnostic performance assessment of EU TIRADS, SWE and microvascularity

A multi-modal assessment of EU TIRADS with the ideal SWE index and the best vascularity mode was performed using the sub-category of nodules with equivocal cytology. The SWE L_{Min} index at a cut-off of 6.1 kPa was combined with qualitative APDI microvascularity assessment and EU TIRADS using category 4 as the optimal cut-off point. This resulted in the best diagnostic performance outcomes in these nodules. The results are shown in Table 5.5. Among different combined approaches, EU TIRADS combined with APDI resulted in the highest overall diagnostic performance (SEN: 88.9%; SPEC: 76.7%; DOR: 26.3; AUROC: 0.83). The multi-modal assessment approach of EU + L_{Min} + $APDI_{qual}$ resulted in a similar sensitivity and significantly improved the specificity and overall diagnostic efficacy of EU alone (SEN: 88.9% for both; SPEC: 66.7% vs 36.7%, $p < 0.05$; AUROC: 0.78 vs 0.62, $p < 0.01$). When compared to the EU alone, the multi-modal approach of combining EU and L_{Min} resulted in a reduction in sensitivity but an improvement in specificity (SEN: 88.9% vs 66.7%, $p > 0.05$; SPEC: 36.7% vs 80%, $p < 0.05$).

Table 5.5: Multi-modal diagnostic performance assessment of EU, SWE and APDI in the stratification of nodules with equivocal cytology

Diagnostic test mode	**Diagnostic performance measures**					
	SEN (%)	**SPEC (%)**	**PPV (%)**	**NPV (%)**	**DOR**	**AUROC**
EU	88.9	36.7	29.6	91.7	4.6	0.62
$APDI_{qual}$	100	76.7**	56.3	100	59.5	0.88***
L_{Min} (kPa)	66.7	80.0** *	50.0	88.9	8.0	0.75
EU + $APDI_{qual}$	88.9	76.7**	53.3	95.8	26.3	0.83**
EU + L_{Min}	66.7	80.0***	50.0	88.9	8.0	0.75
EU + L_{Min} +$APDI_{qual}$	88.9	66.7*	44.4	95.2	16.0	0.78**

EU = European TIRADS, $APDI_{qual}$ = AngioPLUS + Power Doppler Imaging qualitative assessment, L_{Min} = Longitudinal minimum, kPa = kiloPascals
** = p < 0.05; ** = p < 0.01 ; *** = p < 0.001 relative to EU TIRADS*

5.4 Discussion

The present study evaluated the diagnostic value of SWE when combined with EU TIRADS and in multi-modal thyroid nodule assessment involving EU TIRADS, SWE and microvascularity assessment with AngioPLUS. The SWE diagnostic performance outcomes were evaluated based on potential influencing factors such as the imaging scan plane, nodule size, cytology status, and the vascularity status of the nodules.

5.4.1 SWE measurement assessments based on the scan planes

In the present study, the median of the measurements for SWE parameters in the transverse plane were significantly lower than those in the longitudinal plane for the minimum and mean SWE indices in the analysis of all nodules. Similarly, previous studies have suggested that measurements in the longitudinal plane are more likely to result in higher measurement values due to shear waves being propagated faster when the transducer is parallel to the neck muscle fibres [189, 306]. However, our findings differed for the maximum SWE index where the measurements in transverse plane were not significantly different from those in longitudinal plane (43.6 kPa vs 42.3 kPa, $p = 0.56$). This may be explained by the different SWE measurement methodologies particularly the ROI stiffness measurement approach. While our study traced the whole nodule ROI to quantify the stiffness, both previous studies used different approaches of focal circular ROI placement over the visually stiffest nodule area as assessed on the elasticity color map. One of the studies used either 2 or 3 ROIs depending on the size of the nodule [189], while the other study used one ROI [306].

For the malignancy risk stratification of the thyroid nodules, based on the scanning plane used, only the mean SWE index demonstrated statistically significant differences between benign and malignant thyroid nodules using both scanning planes for all nodules, vascular nodules and nodules between 1 to 2 cm in size. Although very few studies have evaluated SWE

measurements concurrently in both planes, our findings concur with two previous studies and they suggested that a good concordance between transverse and longitudinal measurements can be achieved using the mean SWE index [303, 306]. Nonetheless, their conclusions did not consider the influence of nodule size, which was done in the present study. The advantage of a SWE index whose measurements differ distinctly between benign and malignant thyroid nodules regardless of the scan plane used is the ability to still obtain accurate results using a scan plane best suited for a patient's condition or nodule location. For example, imaging a nodule that is very proximal to the pulsating carotid artery in the transverse plane may result in stiffness measurement errors due to motion artifacts, however, these can be avoided by using the longitudinal plane [183, 307]. Therefore, the routine clinical adoption of SWE requires standard cut-off measurement criteria for different scan planes which in turn may be influenced by the size of the nodule among other factors.

5.4.2 Diagnostic performance of SWE indices in combination with EU TIRADS

5.4.2.1 Analysis of all nodules without size stratification

In the present study, the diagnostic performance evaluation of all nodules without size stratification showed that no SWE index performed superior to EU TIRADS. Our findings concurred with those of Swan *et al* [308] for which no SWE index outperformed the French TIRADS (an earlier version of the EU TIRADS) [32] which yielded a sensitivity of 90 % and specificity of 22%. Several studies have evaluated the diagnostic value of the combined SWE and different TIRADS and yielded variable results. Some studies found the addition of SWE to grey scale ultrasound assessment with or without TIRADS to have no diagnostic value, while others found an improved diagnostic performance as evidenced by an increase in sensitivity and/or specificity [190, 304, 305, 309-312]. Similar to Kim *et al*.,[313] the specificities of sole SWE indices in the present study were significantly higher than those of EU TIRADS alone, however the combination of EU TIRADS with SWE indices did not improve the overall diagnostic

performance. Contrarily, a study using French TIRADS, reported high sensitivities in both TIRADS and SWE indices and an improved specificity when the minimum SWE index was combined to the TIRADS [310]. The differences can be attributed to diverse study designs and SWE measurement methods. While the present study used manual tracing of entire nodule for SWE measurement, the previous studies used a fixed ROI and set it at either 2 or 3 mm placed over the stiffest portion of the nodule as determined visually from the elasticity colour scale. Total nodule tracing has been suggested to be more reproducible with good intra- and inter-rater agreement and is thereby more objective than subjective focal nodular ROI placement [314]. Lack of standardization impedes the routine clinical adoption of SWE in thyroid nodule differentiation.

5.4.2.2 Analysis based on size stratifications

The influence of the size of the nodule on combined diagnostic performance outcomes has not been extensively explored. Although some studies concluded that nodule size does not affect shear wave elasticity indices [313, 315], similar to other recent studies, the present study found that nodule size may influence SWE indices [310, 316]. In the present study, no SWE index resulted in a statistically significant difference in the SWE measurements between benign and malignant thyroid nodules < 1 cm. This concurred with other studies that malignant nodules < 1 cm may have low stiffness and SWE may have poor sensitivity in discriminating them from benign nodules [317-319]. In a previous study that used the optimal cut-off of 22.8 kPa for nodules ≤ 1 cm, the mean SWE index demonstrated lower diagnostic performance than for nodules between 1 and 3 cm (SEN: 71.9% vs 82.8%, SPEC: 72.2% vs 83.9% and AUROC 0.73 vs 0.88) [320]. Contrarily, Wang *et al.*,[316] suggested that using the mean SWE index, the combination of SWE and grey scale ultrasound features resulted in improved specificity from that of sole grey scale ultrasound in nodules < 1 cm (SEN: 91.4% to 80.7%, SPEC: 66.7% to 77.8%). Another recent study suggested that the combination of the maximum SWE index at a cut-off of 28.2 kPa with

any suspicious grey scale ultrasound feature had high sensitivity (84.2%) and specificity (75%) in discriminating sub-centimetre nodules [321]. The SWE ROI measurements were all different in the aforementioned studies, with one using 2 mm fixed ROI over the visually stiffest portion [320], another study using a circular Q-box to encompass the whole nodule [316] while the latter was based on perinodular or false rim stiffness [321].

Factors that have been suggested to affect SWE measurements in sub-centimetre nodules include the deep location of such nodules and their pathological nature [25, 310]. Papillary thyroid microcarcinomas (PTMC) have been suggested as less stiff compared to larger classical papillary thyroid carcinomas [320]. Morphological and pathophysiological changes in thyroid cancer involve invasive stromal reactions that alter flexible collagen fibre to stiff collagen, cystic changes, the formation of psammoma bodies and fibrosis which result in the nodule hardening [322, 323]. Therefore, small malignant nodules that lack aggressive fibrosis or calcifications, likely have stiffness changes that are not detected as significant by SWE.

For nodules of sizes between 1 to 2 cm and > 2 cm, the T_{SD} SWE index at cut-offs of 8.7 kPa and 10.7 kPa, respectively, achieved the best diagnostic performance amongst the SWE parameters. The T_{SD} SWE index resulted in an improved specificity with minimal reduction in sensitivity when combined with EU TIRADS. Literature on diagnostic performance outcomes for SWE indices combined with TIRADS based on nodule size stratification is scant. However, in one study the mean SWE index was significantly higher in malignant nodules than benign nodules in similar size stratifications as the present study [316]. The combined assessment of SWE with TIRADS resulted in a higher sensitivity but reduced the specificity in nodules between 1 to 2 cm (SEN: 84.9% to 92.4%; SPEC: 71.4% to 64.9%) and > 2 cm (SEN: 93.2% to 94.3%, SPEC: 62.2% to 53.3%) in the same study. However, the cut-off points at which these diagnostic performance outcomes were not clearly stated. In a recent study, Li et al., [324] reported that the maximum SWE index in the longitudinal plane yielded high sensitivity and

specificity at optimal cut-offs of 37.7 kPa for nodules > 1 cm (SEN: 89.7% and SPEC: 82.6%) and 55.1 kPa for nodules > 2 cm (SEN: 96.4% and SPEC: 88.2%), respectively. However, their study evaluated sole SWE performance without the assessment of combined performance with TIRADS assessment. Contrarily, another study reported high diagnostic performance outcomes using the mean SWE index at a cut-off of 43.3 kPa for nodules between 1 to 3 cm (SEN: 82.8% and SPEC: 83.9%) and 42.7 kPa for nodules > 3 cm (SEN: 72.7% and SPEC: 91.4%) [320]. However, combined assessment with TIRADS was conducted only for all overall nodules but not for the size stratification groups in that study and the results showed no improvement in diagnostic performance.

The challenge of diverse SWE index parameters and optimal cut-off points is well-established in the literature. The threshold setting of the SWE quantitative measurement scale for different studies varies between either 0 to 100 kPa or 0 to 180 kPa which possibly influences the maximum SWE index cut-offs the most [302, 307, 311, 312, 316]. Different SWE measurement methods may explain the variable mean SWE index cut-offs. The focal nodule ROI placement is subjective and does not fully account for the anisotropic nature of soft tissue and nodule elasticity heterogeneity [188, 325]. Some studies have recommended the SD SWE index as a better parameter for measuring the thyroid tissue stiffness heterogeneity than other elasticity indices [188, 308, 326]. Two different studies achieved similar cut-off points but contrasting diagnostic performance outcomes using the SD index for the sole SWE assessment of all nodules using the focal nodule circular ROI placement method (6.5 kPa - SEN:50% and SPEC: 96.6, and 6.8kPa - SEN: 95.5% and SPEC: 42.1%) [188, 310]. The present study established that the combination of EU TIRADS and the SD SWE index using the total nodular ROI tracing method may best discriminate nodules > 1 cm and > 2 cm. As the SD SWE index may best represent heterogeneous fibrotic changes , it is likely more accurate when used using total nodule tracing for stiffness measurements. Since heterogeneous fibrotic changes are usually found in

malignant nodules [322, 327], the detection of these changes by the SD SWE index can help in differentiating benign and malignant nodules. However, fibrotic changes that result in elevated stiffness can also manifest in benign thyroid diseases such as different forms of thyroiditis and calcified multinodular goitres [322, 328].Therefore, the SD SWE index in malignancy risk stratification may be most applicable when there is no suspicion or co-existence of these conditions in focal lesions.

Different thyroid management guidelines use the nodule size as one of the criteria to determine the treatment approach and recommend conservative treatment for indolent sub-centimetre nodules taking into account the patients' preferences [8, 80, 135, 329]. Therefore, we speculate that the complementary use of SWE with specific TIRADS will be beneficial when it is informed by the size of the nodule. Such an approach may limit the overdiagnosis of clinically insignificant sub-centimetre nodules for which active surveillance suffices rather than biopsy or surgery.

5.4.2.3 Analysis of cytologically equivocal nodules

In the present study, the minimum SWE index in the longitudinal plane at the optimal cut-off of 6.1 kPa yielded high specificity (78.4%) with lower sensitivity (60%) with sole SWE analysis in cytologically equivocal thyroid nodules. Similarly, Bardet *et al.* [34] reported a sensitivity of 85% and a specificity of 94% using a 65 kPa cut-off point for the maximum SWE index, while for Samir *et al* [185], a 22.3 kPa cut off point for the mean SWE index had a sensitivity of 82% and specificity of 88%. Contrarily, Chen *et al.* [33] concluded that the SD SWE index at a cut-off of 3.3 kPa had the best diagnostic performance (SEN: 100% and SPEC: 49.8%). However, that study included nodules with non-diagnostic cytology, which are not typically classified as equivocal. Furthermore, another study using a shear wave velocity maximum cut-off of 3.59 m/s achieved a higher sensitivity (83.9%) and a comparable but

slightly lower specificity (79.2%) [330]. The study designs, sample sizes and SWE techniques and ROI measurements varied across the different studies thereby explaining the different study outcomes. Furthermore, while our study and the two other studies [185, 330] only considered cytology categories 3 and 4 as equivocal, Bardet *et al* [34] included the suspicion for malignancy (category 5). We excluded this category because of its inherent high risk of malignancy (60% to 75%) [99], which was 100% in our study upon final histopathology diagnosis.

Recently, EU TIRADS was reported as diagnostically inefficient in the management of follicular neoplasms [149, 291]. Due to sample size limitations, the sub-group analysis of the different equivocal cytology categories was not conducted in the present study. However, the combination of EU TIRADS with SWE improved the overall diagnostic efficacy (0.58 to 0.72) and specificity (37.8% to 78.4%) but lowered the sensitivity (80% to 60%). Nonetheless, the diagnostic performance of different TIRADS in combination with SWE for differentiating cytologically-equivocal nodules has minimal exploration in the literature.

Some previous studies suggested that the combination of strain elastography and/or SWE with grey scale ultrasound features had the potential for improving the diagnostic efficacy in cytologically equivocal nodules [195, 196, 331]. Contrary to our findings, Celletti *et al.* [332] asserted that integration of K-TIRADS with strain elastography resulted in superior diagnostic performance with improved sensitivity (71.4% to 92.9%) but reduced specificity (82.4% to 76.5%). Similarly, the combination of ACR category 5 with the maximum SWE index at a 41.2 kPa cut-off had similar outcomes of raised sensitivity (81.4% to 94.2%) and reduced specificity (84.8% to 75.7%) in another study [333]. ATA TIRADS at a cut-off of category 4 in combination with the shear wave maximum velocity cut off of 3.35 m/s for cytology category 3 nodules, increased both sensitivity and specificity (SEN: 87.5% to 90.6%, and SPEC: 80% to 93.3%) in one study [334]. The different study designs and methodologies explain the differences in

findings. Shear wave elastography techniques varied across the different studies and malignancy risk criteria vary across different TIRADS.

The combination of EU TIRADS with SWE based on the choice of cut-off points in the present study likely contributed to the reduction in the sensitivity. As such, although the overall diagnostic efficiency improved significantly, the loss in sensitivity renders the diagnostic utility of this combined approach vague. It is because substantial reduction of sensitivity leads to increase in false-negative cases and a delay in treatment of cancer patients. Future larger prospective studies are warranted to validate clarify the diagnostic value of this combined approach in cytologically-equivocal nodules.

5.4.2.4 Analysis based on the vascularity status

In the present study, no SWE index outperformed EU TIRADS and there was no improved diagnostic performance with combined approaches in the assessment of all vascular nodules.. The optimal cut-off point for the minimum SWE index was higher for vascular nodules (6.1 kPa) than for avascular nodules (0.88 kPa). Currently, there is limited literature on the evaluation of the influence of the vascularity status of the nodule on SWE elastography in thyroid nodules. However, contrary to the present study, one study using comb-push shear wave elastography reported that the mean SWE index of vascular nodules was lower than that of avascular nodules (27.1 kPa vs 34.8 kPa) [189]. This can be attributed to the differences in sample sizes and elasticity measurement approaches. While the present study used the total nodule trace mode for SWE measurement of the ROI using all indices, the previous study focused on the averages of mean and maximum indices using multiple 3mm circular ROIs over the visual stiffest portion of the nodule. Due to the low sample size of subcategories of thyroid nodule vascularity in the present study, the results are interpreted with caution. Larger

prospective studies are necessary to elucidate the correlation between vascularity and SWE measurements in thyroid nodule differentiation.

5.4.3 Multi-modal assessment of equivocal nodules

In the present study, APDI had superior discrimination ability than SWE for complementary use with EU TIRADS as evidenced by its higher DOR in the sole (59.5 vs 8.0) and combination assessments (26.3 vs 8.0). The combined assessment, of APDI + EU TIRADS resulted in an overall increase in diagnostic efficacy stemming from a sustained high sensitivity with significant improvement in specificity. Similarly, the multi-modal assessment incorporating EU TIRADS, APDI and SWE, maintained high sensitivity, with specificity that was higher than sole EU TIRADS but lower than the EU TIRADS + APDI combination approach.

To the best of my knowledge, currently, no other study has evaluated the diagnostic efficacy of a multi-modal ultrasound approach incorporating EU TIRADS, AngioPLUS power Doppler Imaging and SWE in thyroid nodule assessment. Two studies on multi-modal ultrasound with the integration of superb microvascularity imaging to grey scale ultrasound and strain elastography in diagnosing unequivocal nodules reported conflicting results [199, 296]. Ahn *et al.* [296] concluded that the multi-modal approach improves the specificity of KSThR TIRADS assessment, whereas Yoon *et al*, [199] reported no improvement in the diagnostic performance of using one or more malignancy suspicion grey scale ultrasound features. The differences in the elasticity and microvascularity Doppler measurement techniques likely explain the variances in the studies.

Although limited, studies on the multi-modal ultrasound assessment of cytologically-equivocal thyroid nodules have also shown variable results. A combined approach incorporating strain elastography, grey scale ultrasound, SWE and molecular analysis improved the sensitivity of sole grey scale ultrasound (73.7% to 89.5%) while maintaining an acceptable specificity of

75% [35]. Contrastingly, the combination of grey scale ultrasound, strain elastography and contrast-enhanced ultrasound reduced both sensitivity and specificity outcomes in another study (SEN: 79% to 64% and SPEC: 100% to 92%) [200]. However, in another study both the sensitivity and specificity were improved with the multi-modal assessment of grey scale ultrasound combined with strain elastography and 3D colour Doppler ultrasound (SEN: 57.1% to 85.7% and SPEC: 67.4% to 88.3%) [201]. The varying diagnostic performance outcomes of previous studies can be explained by the diversity of the ultrasound modality combinations whose individual diagnostic efficacies vary.

Although the multi-modal assessment with the three ultrasound imaging modalities in the present study had superior overall diagnostic efficiency than EU TIRADS alone, there was a slight reduction in specificity (66.7%) compared to the integration of EU TIRADS with APDI (76.7%). We alluded this to the criteria for the optimal cut-off points for the different modalities and potential confounding effects affecting the combination of the different modalities. For clinical application considerations, we postulate that the combination of EU TIRADS with APDI may suffice for the diagnosis of cytologically-equivocal thyroid nodules without the inclusion of SWE. Its highest DOR amongst the different combination approaches in this study renders it most diagnostically effective. Nevertheless, there is a need for further validation studies with a larger population sample size based on individual equivocal cytology categories and taking into consideration the potential influence of nodule size on SWE outcomes. Admittedly, the overestimation of outcomes cannot be excluded due to the small sample size in the present study. However, these preliminary findings of the diagnostic performance of multi-modal thyroid ultrasound using EU TIRADS, SWE and APDI confirm the potential of this combined assessment approach for improving the diagnosis and risk-stratification of cytologically-equivocal thyroid nodules by using multiparametric analysis.

5.5 Limitations

There are limitations in the present study. The sample size was small which limited the combined and multi-modal assessment based on the size stratification of the thyroid nodules. Selection bias cannot be excluded due to the selection of patients with FNAC and/or histopathology results. This influenced the high malignancy rate (35.7%) and greater distribution of patients with equivocal cytology thereby limiting the generalisability of the findings to other population contexts. The combination of the equivocal cytology categories 3 and 4 into one group limited the comprehensive assessment of the diagnostic value of SWE in equivocal nodules. Although this approach was adopted due to the limited sample size, the value of multi-modal ultrasound thyroid assessment in equivocal cytology can be inferred from the promising findings in this study and guide future larger prospective studies. Due to the predominance of PTC (86.7%) in this study, the interpretation of the findings may be best applicable to this type of thyroid.

In the present study, only one rater conducted the SWE measurements and intra- and inter-rater reliability assessments were not conducted. However, because the SWE ROI stiffness measurements were conducted for 3 different acquisitions of the same nodule, random measurement errors were limited. The limited assessment of SWE combined with EU TIRADS for various categories of thyroid nodules warrants further large prospective studies with more representation of different types of thyroid cancer and patient demographics. The diagnostic performance assessment of varied SWE ROI measurement techniques relative to the size of the nodules for the same population sample is recommended in future prospective studies to ascertain the most ideal approach.

5.6 Conclusions

This study investigated the diagnostic efficacy of SWE in combination with EU TIRADS for different sub-categories of thyroid nodules, and the multi-modal ultrasound assessment of thyroid nodules with equivocal cytology using EU TIRADS, SWE and APDI. The mean SWE index demonstrated the ability to discriminate benign and malignant thyroid nodules in both longitudinal and transverse planes. The study revealed that the diagnostic performance of SWE in combination with EU TIRADS was influenced by the nodule size having a good diagnostic value in nodules > 1 cm. Although the combination of EU TIRADS and minimum SWE index improved the specificity in equivocal nodules and avascular nodules, the sensitivity was consistently and moderately lower than EU TIRADS alone. The multi-modal assessment demonstrated superior diagnostic performance than EU TIRADS alone in discriminating thyroid nodules with equivocal cytology.

Based on the study results for potential clinical application considerations we posit that using a cut-off of 4 (indeterminate ultrasound suspicion) with EU TIRADS, the addition of SWE using the SD SWE index at cut-offs of 6.8 kPa and 8.8 kPa for nodules > 1 cm and > 2 cm respectively, achieves high specificity. This implies that the combined approach may best rule in disease and limit overtreatment. Stiffness as an additional complementary feature to grey scale ultrasound features in EU TIRADS may be more valuable in thyroid nodules with equivocal grey scale ultrasound features since the presence of any suspicious grey scale ultrasound feature indicates high malignancy risk with this TIRADS. This may best guide the triaging of candidates for follow-up ultrasound or biopsy. Multi-modal assessment using microvascularity techniques, such as APDI, and EU TIRADS would be most beneficial in the follow-up ultrasound of thyroid nodules with equivocal cytology to triage patients scheduled for repeat FNAC.

Chapter 6

Summary of the thesis

Thyroid cancer, the most common endocrine malignancy globally, often presents as nodules which are primarily diagnosed with ultrasound imaging, followed by FNAC where necessary. Ultrasound imaging presents the diagnosis challenge of having no feature that is singularly highly predictive of malignancy. FNAC, the pre-operative reference standard for thyroid cancer diagnosis, has the challenge of up 30% of equivocal cytology results. Although most of the cytologically-equivocal nodules tend to be benign, they require either a repeat biopsy or diagnostic thyroidectomy for a conclusive diagnosis. Thyroid surgery has risks of complications and cost and quality of life implications owing to lifetime hormone replacement therapy after surgery. The advances in ultrasound imaging technology and increased use of FNAC have contributed to an increase in the detection of thyroid cancers particularly sub-clinical cases for which mortality is minimal. There is a need for optimal diagnosis approaches in thyroid cancer to avoid overdiagnosis and the unnecessary invasive treatment of patients.

In grey scale ultrasound feature assessment, different TIRADS emerged to improve the malignancy risk stratification of thyroid nodules by using a combination of multiple features suggestive of malignancy or benign disease to create low-to-high risk categories. Computer-aided diagnosis techniques have also emerged to counter inter- and intra-observer variabilities that arise with the subjective interpretation of sonographic features. AmCAD is the only globally-approved commercial thyroid ultrasound CAD software to incorporate multiple TIRADS for thyroid nodule risk-stratification. Although its diagnostic performance has been evaluated in a few previous studies, the adjustment of different settings for improved accuracy has not yet been explored. Furthermore, the comparative diagnostic performance analysis of the CAD to that of subjective interpreters using the same multiple TIRADS is limited. The first

study (Chapter 3) involved 205 nodules and validated the AmCAD software by comparing the diagnostic performance of 6 TIRADS using the default setting and 3 adjusted settings of margins, anechoic area, and hyperechoic foci. The results of the study confirmed the default setting as the most optimal setting for achieving the highest sensitivity, with EU TIRADS showing the best sensitivity (82.7%). The comparative analysis of CAD with computer-assisted subjective assessment included 162 nodules and 4 TIRADS. The results showed that the sensitivity of the computer-assisted subjective assessment approach was comparable to that of CAD, while the specificity of the CAD was lower for all TIRADS. The diagnostic performance outcomes between EU TIRADS and the KSThR TIRADS were comparable in stratifying all nodules, however, KSThR TIRADS had a slightly lower specificity. For accurately stratifying thyroid nodules, the EU TIRADS suffices for ruling in thyroid cancer regardless of the assessment approach used.

Thyroid ultrasound imaging has evolved to involve advanced Doppler ultrasound imaging techniques. The role of vascularity assessment in aiding the diagnosis of thyroid nodules remains contested. However, it is purported that increased central vascularity of thyroid nodule is suggestive of malignancy while peripheral vascularity suggests benign disease. Thyroid nodule vascularity assessment often involves visual pattern analysis on conventional colour and power Doppler ultrasound images. However, these conventional Doppler ultrasound modes fall short in the detection of low blood flow. AngioPLUS is a novel advanced Doppler ultrasound technique that has high sensitivity and image resolution in detecting low blood flow in microvasculature. The diagnostic value of AngioPLUS in the diagnosis of thyroid nodules lacks investigation. In study two (Chapter 4), the diagnostic performance of AngioPLUS coupled with the conventional colour and power Doppler ultrasound techniques was investigated in comparison with sole conventional Doppler ultrasound techniques. The study included 94 nodules of which 40 had equivocal cytology. The diagnostic performance analyses

were based on visual regional vascularity grading and quantitative RVI analysis. The results demonstrated that the RVI method could exclude benign nodules, however, it was poor at detecting malignant nodules even in combination with EU TIRADS. Visual grading of APDI combined with EU TIRADS improved the specificity of EU TIRADS alone without significantly offsetting the sensitivity (SPEC: 50% to 84.4%, $p < 0.05$; SEN: 83.3% to 76.7%, $p > 0.05$) in the diagnosis of all nodules. The combination of EU TIRADS+APDI had an exceptionally good diagnostic efficacy in the diagnosis of the subcategory of nodules with equivocal cytology, maintaining high sensitivity and significantly improving the specificity from that of EU TIRADS alone (AUROC: 0.62 to 0.89, $p < 0.05$; SEN: remains 88.9%, and SPEC: 38.7% to 77.4%, $p < 0.05$). Since APDI has high diagnostic efficacy in the risk stratification of thyroid nodules, it should be considered in routine thyroid ultrasound imaging to complement EU TIRADS, particularly for cytologically-equivocal nodules.

SWE is among the recent advances in thyroid ultrasound imaging. SWE is a quantitative elastography approach for tissue elasticity measurement. Malignant nodules are often stiffer than benign nodules owing to pathophysiological changes that occur with carcinogenesis. The challenge with the full clinical adoption of SWE in thyroid nodule diagnosis is that past studies have shown variable results over its diagnostic value. Furthermore, there is no consensus regarding the appropriate SWE index, cut-off values, the influence of nodule size, and its role in the diagnosis of nodules with equivocal cytology remains ambiguous. Study three (Chapter 5) investigated the diagnostic value of SWE in combination with EU TIRADS and multimodal assessment for the diagnosis of thyroid nodules. The study included 126 nodules and the diagnostic performance assessment of subcategories of these nodules was conducted based on nodule size, vascularity status and cytology result. The study results showed that the mean SWE index was the only index that could discriminate thyroid nodules in all nodules in both transverse and longitudinal planes. However, no SWE index outperformed EU TIRADS in

stratifying all nodules. The SD SWE index in the transverse plane showed an improvement in the specificity of EU TIRADS while maintaining a slightly lower sensitivity in nodules between 1 and 2 cm at the optimal cut-off of 8.7 kPa (SPEC: 55.9% to 76.5%, $p < 0.05$; SEN: 88.9% to 72.2%, $p > 0.05$) and nodules > 2cm at the optimal cut-off of 10.7 kPa (SPEC: 62.5% to 95.8%, $p < 0.05$; SEN:85.7% to 71.4%, $p > 0.05$). The study findings demonstrated that there is a need to consider nodule size in the routine SWE ultrasound assessment of thyroid nodules.

The role of SWE in cytologically equivocal nodules and avascular nodules requires further exploration. However, the combination of EU TIRADS with the minimum SWE index in the longitudinal plane at optimal cut-offs of 6.1 kPa and 0.9 kPa respectively for the two groups of nodules, improved the specificity and the overall diagnostic efficacy but significantly lowered the sensitivity. The diagnostic efficiency of these cut-off points should be evaluated further and validated in future larger prospective studies. In comparison with EU TIRADS, a multimodal approach of EU TIRADS, ADPI and SWE retained a high sensitivity (88.9%) and improved the specificity (36.7% to 66.7%, $p < 0.05$) in diagnosing cytologically equivocal nodules. However, its specificity and DOR were lower than those of the combination of EU TIRADS and ADPI. Therefore, the combination of EU TIRADS and ADPI is ideal for the diagnosis of thyroid nodules with equivocal cytology and is recommended for implementation in the routine ultrasound evaluation of these nodules.

The study findings in this thesis have clinical significance and potential value for clinical application, however; challenges persist in view of clinical implementation. Firstly, since different clinical settings use varied thyroid diagnosis and management guidelines, there is need for effective collaboration with clinicians to adopt the most efficient ultrasound approaches for thyroid nodule diagnosis. A challenge remains in determining the TIRADS that best optimises both sensitivity and specificity, although the EU and KSThR TIRADS could best rule in malignancy based on CAD and subjective approaches in the present study. In

addition, while the AngioPLUS technique demonstrated the best diagnostic efficacy in thyroid nodule diagnosis, the major challenge with its routine clinical adoption is likely that the software is vendor-specific. AngioPLUS is only found in Supersonic Aixplorer ultrasound machines, yet clinicians may have their own preferences of ultrasound machine. However; meaningful diagnostic utility research findings as shown in the present study may potentially guide the choice of ultrasound machines that may best optimise patient diagnosis. SWE is another vendor-specific mode which showed potential for the optimised diagnosis of thyroid nodules > 1cm. However, the major challenge with the routine clinical adoption of this technique is the lack of a standard ROI measurement approach that best results in improved diagnostic efficacy. The manual outlining method and the standard deviation parameter showed great potential in the present study, however; further large prospective studies are warranted to validate our findings and ascertain the best approach, particularly when factoring in the size of the nodules.

In summary, this thesis investigated the diagnostic value of computer-aided diagnosis, AngioPLUS coupled with Doppler ultrasound and SWE in the differentiation of thyroid nodules. The thesis concludes with the following considerations for clinical application:

1. The key role of CAD in screening patients with thyroid nodules for cancer suspicion is affirmed, however, subjective interpretation remains necessary more so in the use of specific TIRADS since it results in higher specificity than CAD. CAD should complement subjective interpretation, serving as a faster tool for triaging patients.
2. Microvascularity assessment with AngioPLUS coupled with power Doppler ultrasound based on qualitative regional vascularity grading has immense potential for improving the accuracy of the EU TIRADS in the stratification of thyroid nodules, more so in nodules with equivocal cytology. This combined approach is advisable in

cytologically equivocal nodules as a follow-up ultrasound procedure and may potentially limit unwarranted biopsies.

3. SWE assessment should take into account the size of nodules and is recommended for nodules ≥ 1cm to improve the diagnostic efficacy of EU TIRADS.
4. Since a combined approach of EU TIRADS and APDI is more diagnostically efficient for diagnosing cytologically-equivocal thyroid nodules than a multimodal approach involving SWE, there may be no need for further SWE assessment when EU TIRADS and APDI both indicate suspicion of malignancy. A more diagnostically-efficient sequence in thyroid ultrasound imaging can involve prioritising grey scale ultrasound assessment and microvascularity assessment over SWE assessment.

References

1. Pellegriti, G.; Frasca, F.; Regalbuto, C.; Squatrito, S.; Vigneri, R., Worldwide increasing incidence of thyroid cancer: Update on epidemiology and risk factors. *Journal of Cancer Epidemiology* **2013,** *2013*, 965212-965212.

2. Zhuang, Y.; Li, C.; Hua, Z.; Chen, K.; Lin, J. L., A novel TIRADS of US classification. *BioMedical Engineering Online* **2018,** *17* (1), 82.

3. Hong_Kong_Cancer_Registry Overview of Hong Kong Cancer Statistics of 2018. https://www3.ha.org.hk/cancereg/allagesresult.asp.

4. Brito, J. P.; Gionfriddo, M. R.; Al Nofal, A.; Boehmer, K. R.; Leppin, A. L.; Reading, C.; Callstrom, M.; Elraiyah, T. A.; Prokop, L. J.; Stan, M. N.; Murad, M. H.; Morris, J. C.; Montori, V. M., The Accuracy of Thyroid Nodule Ultrasound to Predict Thyroid Cancer: Systematic Review and Meta-Analysis. *The Journal of Clinical Endocrinology & Metabolism* **2014,** *99* (4), 1253-1263.

5. Hoang, J. K.; Nguyen, X. V.; Davies, L., Overdiagnosis of Thyroid Cancer. *Academic Radiology* **2015,** *22* (8), 1024-1029.

6. Liu, X.; Medici, M.; Kwong, N.; Angell, T. E.; Marqusee, E.; Kim, M. I.; Larsen, P. R.; Cho, N. L.; Nehs, M. A.; Ruan, D. T.; Gawande, A.; Moore, F., Jr.; Barletta, J.; Krane, J. F.; Cibas, E. S.; Yang, T.; Alexander, E. K., Bethesda Categorization of Thyroid Nodule Cytology and Prediction of Thyroid Cancer Type and Prognosis. *Thyroid* **2016,** *26* (2), 256-61.

7. Bajaj, Y.; De, M.; Thompson, A., Fine needle aspiration cytology in diagnosis and management of thyroid disease. *J Laryngol Otol* **2006,** *120* (6), 467-9.

8. Durante, C.; Grani, G.; Lamartina, L.; Filetti, S.; Mandel, S. J.; Cooper, D. S., The diagnosis and management of thyroid nodules a review. *JAMA - Journal of the American Medical Association* **2018,** *319* (9), 919-924.

9. Bastin, S.; Bolland, M. J.; Croxson, M. S., Role of ultrasound in the assessment of nodular thyroid disease. *Journal of medical imaging and radiation oncology* **2009,** *53* (2), 177-87.

10. Baig, F. N.; Liu, S. Y. W.; Yip, S. P.; Law, H. K. W.; Ying, M. T. C., Update on Ultrasound Diagnosis for Thyroid Cancer. *Hong Kong Journal of Radiology* **2018,** *21*, 82-93.

11. Liang, X.-W.; Cai, Y.-Y.; Yu, J.-S.; Liao, J.-Y.; Chen, Z.-Y., Update on thyroid ultrasound. *Chinese Medical Journal* **2019,** *132* (16), 1974-1982.

12. Hoang, J. K.; Lee, W. K.; Lee, M.; Johnson, D.; Farrell, S., US Features of Thyroid Malignancy: Pearls and Pitfalls. *RadioGraphics* **2007,** *27* (3), 847-860.

13. Moon, W. J.; Jung, S. L.; Lee, J. H.; Na, D. G.; Baek, J. H.; Lee, Y. H.; Kim, J.; Kim, H. S.; Byun, J. S.; Lee, D. H.; Grp, T. S.; Radiolo, K. S. N.-H. N., Benign and malignant thyroid nodules: US differentiation - Multicenter retrospective study. *Radiology* **2008,** *247* (3), 762-770.

14. Remonti, L. R.; Kramer, C. K.; Leitão, C. B.; Pinto, L. C. F.; Gross, J. L., Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid : official journal of the American Thyroid Association* **2015,** *25* (5), 538-550.

15. Campanella, P.; Ianni, F.; Rota, C. A.; Corsello, S. M.; Pontecorvi, A., DIAGNOSIS IN ENDOCRINOLOGY: Quantification of cancer risk of each clinical and ultrasonographic suspicious feature of thyroid nodules: a systematic review and meta-analysis. **2014,** *170* (5), R203.

16. Carmeliet, P.; Jain, R. K., Angiogenesis in cancer and other diseases. *Nature* **2000,** *407* (6801), 249-257.

17. Ringel, M. D.; Burgun, S. J., Thyroid Cancer. 2016; pp 119-129.

18. Rios, A.; Torregrosa; Rodríguez, J. M.; Rodríguez; Cepero; Abellán, M. D.; Torregrosa, N. M.; Hernández; A, M.; Parrilla; Torregrosa, B.; Rodríguez, J. M.; Rodríguez, D.; Cepero, A.; Abellán, M. D.; Torregrosa, N. M.; Hernández, A. M.; Parrilla, P., Ultrasonographic risk factors of malignancy in thyroid nodules. *Langenbeck's Archives of Surgery* **2016,** *401* (6), 839-849.

19. Anil, G.; Hegde, A.; Chong, F. H. V., Thyroid nodules: Risk stratification for malignancy with ultrasound and guided biopsy. *Cancer Imaging* **2011,** *11* (1), 209-223.

20. Xie, C.; Cox, P.; Taylor, N.; LaPorte, S., Ultrasonography of thyroid nodules: a pictorial review. 2016; Vol. 7, pp 77-86.

21. Bercoff, J.; Tanter, M., ULTRASOUND IMAGING GOES ULTRAFAST: A CHANGE IN PARADIGM IN MEDICAL ULTRASOUND. *MEDICAL PHYSICS INTERNATIONAL* **2015,** *3* (2), 109-119.

22. Liu, H.; Liao, Q.; Wang, Y.; Hu, Y.; Zhu, Q. L.; Wang, L.; Liu, Q. Q.; Li, J. C.; Jiang, Y. X., A new tool for diagnosing parathyroid lesions: angio plus ultrasound imaging. *J. Thorac. Dis.* **2019,** *11* (11), 4829-4834.

23. Baig, F. N.; Van Lunenburg, J. T. J.; Liu, S. Y. W.; Yip, S. P.; Law, H. K. W.; Ying, M., Computer-aided assessment of regional vascularity of thyroid nodules for prediction of malignancy. *Scientific Reports* **2017,** *7* (14350), 1-9.

24. Zaleska-Dorobisz, U.; Kaczorowski, K.; Pawlus, A.; Puchalska, A.; Inglot, M., Ultrasound Elastography- Review of Techniques and its clinical applications. *Advanced Clinical Exp Med* **2014,** *23* (4), 645-655.

25. Sebag, F.; Vaillant-Lombard, J.; Berbis, J.; Griset, V.; Henry, J. F.; Petit, P.; Oliver, C., Shear Wave Elastography: A New Ultrasound Imaging Mode for the Differential Diagnosis of Benign and Malignant Thyroid Nodules. *The Journal of Clinical Endocrinology & Metabolism* **2010,** *95* (12), 5281-5288.

26. Zhao, C.-K.; Xu, H.-X., Ultrasound elastography of the thyroid: principles and current status. *Ultrasonography* **2019,** *38* (2), 106-124.

27. Carneiro-Pla, D., Ultrasound elastography in the evaluation of thyroid nodules for thyroid cancer. 2013; Vol. 25, pp 1-5.

28. Hang, J.; Li, F.; Qiao, X. h.; Ye, X. h.; Li, A.; Du, L. f., Combination of maximum shear wave elasticity modulus and TIRADS improves the diagnostic specificity in characterizing thyroid nodules: A retrospective study. *International Journal of Endocrinology* **2018,** *2018*, 1-8.

29. Kwak, J. Y.; Kim, E.-K., Ultrasound elastography for thyroid nodules: recent advances. *Ultrasonography (Seoul, Korea)* **2014,** *33* (2), 75-82.

30. Koh, J.; Moon, H. J.; Park, J. S.; Kim, S. J.; Kim, H. Y.; Kim, E. K.; Kwak, J. Y., Variability in Interpretation of Ultrasound Elastography and Gray-Scale Ultrasound in Assessing Thyroid Nodules. *Ultrasound in Medicine and Biology* **2016,** *42* (1), 51-59.

31. Chang, N.; Zhang, X.; Wan, W.; Zhang, C.; Zhang, X., The Preciseness in Diagnosing Thyroid Malignant Nodules Using Shear-Wave Elastography. *Medical science monitor : international medical journal of experimental and clinical research* **2018,** *24*, 671-677.

32. Russ, G., Risk stratification of thyroid nodules on ultrasonography with the French TI-RADS: description and reflections What Is Risk Stratification and Why and How Should We Use It for Thyroid Nodules? *Ultrasonography* **2016,** *35* (1), 25-38.

33. Chen, L.; Shi, Y.-x.; Liu, Y.-c.; Zhan, J.; Diao, X.-h.; Chen, Y.; Zhan, W.-w., The values of shear wave elastography in avoiding repeat fine-needle aspiration for thyroid nodules with nondiagnostic and undetermined cytology. *Clinical Endocrinology* **2019,** *91* (1), 201-208.

34. Bardet, S.; Ciappuccini, R.; Pellot-Barakat, C.; Monpeyssen, H.; Michels, J.-J.; Tissier, F.; Blanchard, D.; Menegaux, F.; de Raucourt, D.; Lefort, M.; Reznik, Y.; Rouxel, A.; Heutte, N.; Brenac, F.; Leconte, A.; Buffet, C.; Clarisse, B.; Leenhardt, L., Shear Wave

Elastography in Thyroid Nodules with Indeterminate Cytology: Results of a Prospective Bicentric Study. *Thyroid* **2017,** *27* (11), 1441-1449.

35. Gay, S.; Schiaffino, S.; Santamorena, G.; Massa, B.; Ansaldo, G.; Turtulici, G.; Giusti, M.; At The Policlinico San Martino Genoa, T. T.; Genoa, Role of Strain Elastography and Shear-Wave Elastography in a Multiparametric Clinical Approach to Indeterminate Cytology Thyroid Nodules. *Medical science monitor : international medical journal of experimental and clinical research* **2018,** *24*, 6273-6279.

36. Stathatos, N., Anatomy and Physiology of the Thyroid Gland: Clinical Correlates to Thyroid Cancer. Wartofsky, L.; Van Nostrand, D., Eds. Springer Science+ Business Media: New York, 2016; pp 3-8.

37. Hadjisavva, I. S.; Economides, P. A., Thyroid ultrasound. Linos, D.; Chung, W. Y., Eds. Springer- Verlag Berlin Heidelberg: 2012; pp 17-35.

38. Dagan, R.; Amdur, R. J., Thyroid Cancer. Third ed.; Small, W., Jr., Ed. John Wiley & Sons, Inc: 2017; pp 318-333.

39. Kondo, T.; Ezzat, S.; Asa, S. L., Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nature Reviews Cancer* **2006,** *6* (4), 292-306.

40. Ravera, S.; Reyna-Neyra, A.; Ferrandino, G.; Amzel, L. M.; Carrasco, N., The Sodium/Iodide Symporter (NIS): Molecular Physiology and Preclinical and Clinical Applications. *Annual review of physiology* **2017,** *79*, 261-289.

41. Tate, P., Endocrine System: Thyroid Gland. 2nd ed.; McGrawHill: 2012; pp 441-446.

42. Aw, T. C.; Yap, C. Y. F. *Thyroid Function Tests*; 2011.

43. Doherty, G. M., Thyroid Nodules and Cancer Risk: Surgical Management. Wartofsky, L.; Van Nostrand, D., Eds. Springer Science: New York, 2016; pp 331-333.

44. Lawrence, W.; Kaplan, B. J., Diagnosis and management of patients with thyroid nodules. *Journal of Surgical Oncology* **2002,** *80* (3), 157-170.

45. Singh Ospina, N.; Iñiguez-Ariza, N. M.; Castro, M. R., Thyroid nodules: diagnostic evaluation based on thyroid cancer risk assessment. *BMJ* **2020,** *368*, l6670.

46. Davies, L.; Welch, H. G., Increasing Incidence of Thyroid Cancer in the United States, 1973-2002. *JAMA* **2006,** *295* (18), 2164-2167.

47. Valderrabano, P.; McIver, B., Evaluation and Management of Indeterminate Thyroid Nodules: The Revolution of Risk Stratification Beyond Cytological Diagnosis. 2017; Vol. 24.

48. Cramer, J. D.; Fu, P.; Harth, K. C.; Margevicius, S.; Wilhelm, S. M., Analysis of the rising incidence of thyroid cancer using the Surveillance, Epidemiology and End Results national cancer data registry. *Surgery* **2010,** *148* (6), 1147-1153.

49. Kitahara, C. M.; Sosa, J. A., The changing incidence of thyroid cancer. *Nature Reviews Endocrinology* **2016,** *12* (11), 646-653.

50. Cui, Y.; Mubarik, S.; Li, R.; Nawsherwan; Yu, C., Trend dynamics of thyroid cancer incidence among China and the U.S. adult population from 1990 to 2017: a joinpoint and age-period-cohort analysis. *BMC Public Health* **2021,** *21* (1), 624.

51. Segkos, K.; Porter, K.; Senter, L.; Ringel, M. D.; Nabhan, F. A., Neck Ultrasound in Patients with Follicular Thyroid Carcinoma. *Hormones and Cancer* **2018,** *1*.

52. Miranda-Filho, A.; Lortet-Tieulent, J.; Bray, F.; Cao, B.; Franceschi, S.; Vaccarella, S.; Dal Maso, L., Thyroid cancer incidence trends by histology in 25 countries: a population-based study. *The Lancet Diabetes & Endocrinology* **2021,** *9* (4), 225-234.

53. Liebner, D. A.; Shah, M. H., Thyroid cancer: Pathogenesis and targeted therapy. 2011; Vol. 2, pp 173-195.

54. Cabanillas, M. E.; McFadden, D. G.; Durante, C., Thyroid cancer. *The Lancet* **2016,** *388* (10061), 2783-2795.

55. Kebebew, E.; Greenspan, F. S.; Clark, O. H.; Woeber, K. A.; McMillan, A., Anaplastic thyroid carcinoma. *Cancer* **2005,** *103* (7), 1330-1335.

56. Hahn, S. Y.; Shin, J. H., Description and comparison of the sonographic characteristics of poorly differentiated thyroid carcinoma and anaplastic thyroid carcinoma. *Journal of Ultrasound in Medicine* **2016,** *35* (9), 1873-1879.

57. DeLellis, R. A., Pathology and genetics of thyroid carcinoma. *Journal of Surgical Oncology* **2006,** *94* (8), 662-669.

58. Doherty, G. M., Thyroid Cancer. 2016; pp 779-780.

59. Nikiforova, M. N.; Kimura, E. T.; Gandhi, M.; Biddinger, P. W.; Knauf, J. A.; Basolo, F.; Zhu, Z.; Giannini, R.; Salvatore, G.; Fusco, A.; Santoro, M.; Fagin, J. A.; Nikiforov, Y. E., BRAF Mutations in Thyroid Tumors Are Restricted to Papillary Carcinomas and Anaplastic or Poorly Differentiated Carcinomas Arising from Papillary Carcinomas. *The Journal of Clinical Endocrinology & Metabolism* **2003,** *88* (11), 5399-5404.

60. Fagin, J. A., Genetics of papillary thyroid cancer initiation: implications for therapy. *Trans Am Clin Climatol Assoc* **2005,** *116*, 259-271.

61. Nikiforov, Y. E.; Nikiforova, M. N., Molecular genetics and diagnosis of thyroid cancer. *Nature Reviews Endocrinology* **2011,** *7* (10), 569-580.

62. Xing, M.; Alzahrani, A. S.; Carson, K. A.; Viola, D.; Elisei, R.; Bendlova, B.; Yip, L.; Mian, C.; Vianello, F.; Tuttle, R. M.; Robenshtok, E.; Fagin, J. A.; Puxeddu, E.; Fugazzola, L.; Czarniecka, A.; Jarzab, B.; O'Neill, C. J.; Sywak, M. S.; Lam, A. K.; Riesco-Eizaguirre, G.; Santisteban, P.; Nakayama, H.; Tufano, R. P.; Pai, S. I.; Zeiger, M. A.; Westra, W. H.; Clark, D. P.; Clifton-Bligh, R.; Sidransky, D.; Ladenson, P. W.; Sykorova, V., Association between BRAF V600E mutation and mortality in patients with papillary thyroid cancer. *JAMA* **2013,** *309* (14), 1493-1501.

63. Xing, M.; Westra, W. H.; Tufano, R. P.; Cohen, Y.; Rosenbaum, E.; Rhoden, K. J.; Carson, K. A.; Vasko, V.; Larin, A.; Tallini, G.; Tolaney, S.; Holt, E. H.; Hui, P.; Umbricht, C. B.; Basaria, S.; Ewertz, M.; Tufaro, A. P.; Califano, J. A.; Ringel, M. D.; Zeiger, M. A.;

Sidransky, D.; Ladenson, P. W., BRAF mutation predicts a poorer clinical prognosis for papillary thyroid cancer. *J Clin Endocrinol Metab* **2005,** *90* (12), 6373-9.

64. Kim, T. H.; Park, Y. J.; Lim, J. A.; Ahn, H. Y.; Lee, E. K.; Lee, Y. J.; Kim, K. W.; Hahn, S. K.; Youn, Y. K.; Kim, K. H.; Cho, B. Y.; Park, D. J., The association of the BRAF(V600E) mutation with prognostic factors and poor clinical outcome in papillary thyroid cancer: a meta-analysis. *Cancer* **2012,** *118* (7), 1764-73.

65. Xing, M., BRAF mutation in papillary thyroid cancer: pathogenic role, molecular bases, and clinical implications. *Endocr Rev* **2007,** *28* (7), 742-62.

66. Riesco-Eizaguirre, G.; Rodríguez, I.; De la Vieja, A.; Costamagna, E.; Carrasco, N.; Nistal, M.; Santisteban, P., The BRAFV600E Oncogene Induces Transforming Growth Factor β Secretion Leading to Sodium Iodide Symporter Repression and Increased Malignancy in Thyroid Cancer. *Cancer Research* **2009,** *69* (21), 8317-8325.

67. Russo, M.; Malandrino, P.; Nicolosi, M. L.; Manusia, M.; Marturano, I.; Trovato, M. A.; Pellegriti, G.; Frasca, F.; Vigneri, R., The BRAF(V600E) mutation influences the short- and medium-term outcomes of classic papillary thyroid cancer, but is not an independent predictor of unfavorable outcome. *Thyroid* **2014,** *24* (8), 1267-74.

68. Popoveniuc, G.; Jonklaas, J., Thyroid Nodules. *Med Clin North Am* **2012,** *96* (2), 329-349.

69. Perros, P.; Boelaert, K.; Colley, S.; Evans, C.; Evans, R. M.; Gerrard BA, G.; Gilbert, J.; Harrison, B.; Johnson, S. J.; Giles, T. E.; Moss, L.; Lewington, V.; Newbold, K.; Taylor, J.; Thakker, R. V.; Watkinson, J.; Williams, G. R., Guidelines for the management of thyroid cancer. *Clinical Endocrinology* **2014,** *81* (s1), 1-122.

70. Gharib, H.; Garber, J. R.; Duick, D. S.; Mack Harrell, R.; Hegedüs, L.; Paschke, R.; Valcavi, R.; Vitti, P.; Daniel; Duick, S.; Harrell; Mack, R.; Hegedüs, L.; Paschke, R.; Valcavi, R.; Vitti, P., AACE/ACE/AME Guidelines AMERICAN ASSOCIATION OF

CLINICAL ENDOCRINOLOGISTS, AMERICAN COLLEGE OF ENDOCRINOLOGY, AND ASSOCIAZIONE MEDICI ENDOCRINOLOGI MEDICAL GUIDELINES FOR CLINICAL PRACTICE FOR THE DIAGNOSIS AND MANAGEMENT OF THYROID NODULES-2016 UPDATE AP. *ENDOCRINE PRACTICE* **2016,** *22* (5), 622-639.

71. 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., 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* **2016,** *26* (1), 1-133.

72. Bomeli, S. R.; Lebeau, S. O.; Ferris, R. L.; Md, S. R. B. O. L.; Md, S. R. B. O. L., Evaluation of a thyroid nodule. *Otolaryngologic Clinics of NA* **2010,** *43* (2), 229-238.

73. Raza, S. N.; Shah, M. D.; Palme, C. E.; Hall, F. T.; Eski, S.; Freeman, J. L., Risk factors for well-differentiated thyroid carcinoma in patients with thyroid nodular disease. *Otolaryngology - Head and Neck Surgery* **2008,** *139* (1), 21-26.

74. Gilliland, F. D.; Hunt, W. C.; Morris, D. M.; Key, C. R., Prognostic factors for thyroid carcinoma: A population-based study of 15,698 cases from the Surveillance, Epidemiology and End Results (SEER) program 1973-1991. *Cancer* **1997,** *79* (3), 564-573.

75. Rago, T.; Fiore, E.; Scutari, M.; Santini, F.; Di Coscio, G.; Romani, R.; Piaggi, P.; Ugolini, C.; Basolo, F.; Miccoli, P.; Pinchera, A.; Vitti, P., Male sex, single nodularity, and young age are associated with the risk of finding a papillary thyroid cancer on fine-needle aspiration cytology in a large series of patients with nodular thyroid disease. *European Journal of Endocrinology* **2010,** *162* (4), 763-770.

76. Pathak, K. A.; Mazurat, A.; Lambert, P.; Klonisch, T.; Nason, R. W., Prognostic Nomograms To Predict Oncological Outcome of Thyroid Cancers. *The Journal of Clinical Endocrinology & Metabolism* **2013,** *98* (12), 4768-4775.

77. Haugen, B. R.; Alexander, E. K.; Bible, K. C., 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* **2016,** *26*, 1.

78. Mazurat, A.; Torroni, A.; Hendrickson-Rebizant, J.; Benning, H.; Nason, R. W.; Pathak, K. A., The age factor in survival of a population cohort of well-differentiated thyroid cancer. *Endocrine connections* **2013,** *2* (3), 154-60.

79. Brown, R. L.; de Souza, J. A.; Cohen, E. E., Thyroid cancer: Burden of illness and management of disease. 2011; Vol. 2, pp 193-199.

80. Tamhane, S.; Gharib, H., Thyroid nodule update on diagnosis and management. *Clinical Diabetes and Endocrinology* **2016,** *2* (1), 17-17.

81. Cooper, D. S.; Doherty, G. M.; Haugen, B. R.; Kloos, R. T.; Lee, S. L.; Mandel, S. J.; Mazzaferri, E. L.; McIver, B.; Pacini, F.; Schlumberger, M.; Sherman, S. I.; Steward, D. L.; Tuttle, R. M., Revised American Thyroid Association Management Guidelines for Patients with Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* **2009,** *19* (11), 1167-1214.

82. Al Dawish, M. A.; Alwin Robert, A.; Thabet, M. A.; Braham, R., Thyroid Nodule Management: Thyroid-Stimulating Hormone, Ultrasound, and Cytological Classification System for Predicting Malignancy. *Cancer Inform* **2018,** *17*, 1176935118765132.

83. Boelaert, K., The association between serum TSH concentration and thyroid cancer. *Endocrine-Related Cancer* **2009,** *16* (4), 1065-1072.

84. Nieto, H.; Boelaert, K., WOMEN IN CANCER THEMATIC REVIEW: Thyroid-stimulating hormone in thyroid cancer: does it matter? *Endocrine-Related Cancer* **2016,** *23* (11), T109-T121.

85. Zheng, J.; Li, C.; Lu, W.; Wang, C.; Ai, Z., Quantitative assessment of preoperative serum thyrotropin level and thyroid cancer. *Oncotarget* **2016,** *7* (23).

86. Patel, A.; Shostrom, V.; Treude, K.; Lydiatt, W.; Smith, R.; Goldner, W., Serum Thyroglobulin: Preoperative Levels and Factors Affecting Postoperative Optimal Timing following Total Thyroidectomy. *Int J Endocrinol* **2019,** *2019*, 1384651.

87. Wang, W.; Chang, J.; Jia, B.; Liu, J., The Blood Biomarkers of Thyroid Cancer. *Cancer Manag Res* **2020,** *12*, 5431-5438.

88. Harvey, A., Management of Distant Metastases in Differentiated Thyroid Cancer. Pathak, A. K.; Nason, R. W.; Pasieka, J. L., Eds. Springer India Limited: New Dehli, 2015; pp 61-77.

89. Mitchell, A. L.; Gandhi, A.; Scott-Coombes, D.; Perros, P., Management of thyroid cancer: United Kingdom National Multidisciplinary Guidelines. *The Journal of Laryngology & Otology* **2016,** *130* (S2), S150-S160.

90. Hsiao, S. J.; Nikiforov, Y. E., Molecular approaches to thyroid cancer diagnosis. *Endocrine-Related Cancer* **2014,** *21* (5), T301-T313.

91. Alexander, E. K.; Kennedy, G. C.; Baloch, Z. W.; Cibas, E. S.; Chudova, D.; Diggans, J.; Friedman, L.; Kloos, R. T.; LiVolsi, V. A.; Mandel, S. J.; Raab, S. S.; Rosai, J.; Steward, D. L.; Walsh, P. S.; Wilde, J. I.; Zeiger, M. A.; Lanman, R. B.; Haugen, B. R., Preoperative Diagnosis of Benign Thyroid Nodules with Indeterminate Cytology. *New England Journal of Medicine* **2012,** *367* (8), 705-715.

92. Ulusoy, B., The Management of Thyroid Nodules. *Turk Otolarengoloji Arsivi/Turkish Archives of Otolaryngology* **2016,** *53* (4), 173-182.

93. Conrad, R.; Yang, S. E.; Chang, S.; Bhasin, M.; Sullivan, P. S.; Moatamed, N. A.; Lu, D. Y., Comparison of Cytopathologist-Performed Ultrasound-Guided Fine-Needle Aspiration With Cytopathologist-Performed Palpation-Guided Fine-Needle Aspiration: A Single Institutional Experience. *Arch Pathol Lab Med* **2018,** *142* (10), 1260-1267.

94. Poller, D. N.; Baloch, Z. W.; Fadda, G.; Johnson, S. J.; Bongiovanni, M.; Pontecorvi, A.; Cochand-Priollet, B., Thyroid FNA: New classifications and new interpretations. *Cancer Cytopathol* **2016,** *124* (7), 457-66.

95. Bongiovanni, M., International Comparison Study of Thyroid Reporting Systems. In *Thyroid FNA Cytology: Differential Diagnoses and Pitfalls*, Kakudo, K., Ed. Springer Singapore: Singapore, 2019; pp 49-52.

96. Cibas, E. S.; Ali, S. Z., The Bethesda System for Reporting Thyroid Cytopathology. *American Journal of Clinical Pathology* **2009,** *132* (5), 658-665.

97. Cibas, E. S.; Ali, S. Z., The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Journal of the American Society of Cytopathology* **2017,** *6* (6), 217-222.

98. Bongiovanni, M.; Spitale, A.; Faquin, W. C.; Mazzucchelli, L.; Baloch, Z. W., The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. *Acta cytologica* **2012,** *56* (4), 333-9.

99. Bongiovanni, M.; Bellevicine, C.; Troncone, G.; Sykiotis, G. P., Approach to cytological indeterminate thyroid nodules. *Gland Surg* **2019,** *8* (Suppl 2), S98-S104.

100. Habib, L. A. M.; Abdrabou, A. M.; Geneidi, E. A. S.; Sultan, Y. M., Role of ultrasound elastography in as sessment of indeterminate thyroid nodules. *Egyptian Journal of Radiology and Nuclear Medicine* **2016,** *47* (1), 141-147.

101. Kakudo, K.; El-Naggar, A. K.; Hodak, S. P.; Khanafshar, E.; Nikiforov, Y. E.; Nosé, V.; Thompson, L. D. R., Noninvasive follicular thyroid neoplasm with papillary-like nuclear

features (NIFTP) in thyroid tumor classification. *Pathology International* **2018,** *68* (6), 327-333.

102. Trimboli, P.; Treglia, G.; Guidobaldi, L.; Saggiorato, E.; Nigri, G.; Crescenzi, A.; Romanelli, F.; Orlandi, F.; Valabrega, S.; Sadeghi, R.; Giovanella, L., Clinical characteristics as predictors of malignancy in patients with indeterminate thyroid cytology: a meta-analysis. *Endocrine* **2014,** *46* (1), 52-9.

103. Conzo, G.; Avenia, N.; Ansaldo, G. L.; Calò, P.; De Palma, M.; Dobrinja, C.; Docimo, G.; Gambardella, C.; Grasso, M.; Lombardi, C. P.; Pelizzo, M. R.; Pezzolla, A.; Pezzullo, L.; Piccoli, M.; Rosato, L.; Siciliano, G.; Spiezia, S.; Tartaglia, E.; Tartaglia, F.; Testini, M.; Troncone, G.; Signoriello, G., Surgical treatment of thyroid follicular neoplasms: results of a retrospective analysis of a large clinical series. *Endocrine* **2017,** *55* (2), 530-538.

104. Vuong, H. G.; Ngo, H. T. T.; Bychkov, A.; Jung, C. K.; Vu, T. H.; Lu, K. B.; Kakudo, K.; Kondo, T., Differences in surgical resection rate and risk of malignancy in thyroid cytopathology practice between Western and Asian countries: A systematic review and meta-analysis. *Cancer Cytopathol* **2020,** *128* (4), 238-249.

105. Kakudo, K., Asian and Western practice in thyroid pathology: similarities and differences. *Gland surgery* **2020,** *9* (5), 1614-1627.

106. Kakudo, K.; Higuchi, M.; Hirokawa, M.; Satoh, S.; Jung, C. K.; Bychkov, A., Thyroid FNA cytology in Asian practice-Active surveillance for indeterminate thyroid nodules reduces overtreatment of thyroid carcinomas. *Cytopathology* **2017,** *28* (6), 455-466.

107. Yoon, J. H.; Moon, H. J.; Kim, E. K.; Kwak, J. Y., Inadequate cytology in thyroid nodules: should we repeat aspiration or follow-up? *Ann Surg Oncol* **2011,** *18* (5), 1282-9.

108. Tee, Y. Y.; Lowe, A. J.; Brand, C. A.; Judson, R. T., Fine-needle aspiration may miss a third of all malignancy in palpable thyroid nodules: A comprehensive literature review. 2007; Vol. 246, pp 714-720.

109. Sharma, C., Diagnostic accuracy of fine needle aspiration cytology of thyroid and evaluation of discordant cases. *J Egypt Natl Canc Inst* **2015,** *27* (3), 147-53.

110. Cramer, H., Fine-needle aspiration cytology of the thyroid: an appraisal. *Cancer* **2000,** *90* (6), 325-9.

111. Gul, K.; Ersoy, R.; Dirikoc, A.; Korukluoglu, B.; Ersoy, P. E.; Aydin, R.; Ugras, S. N.; Belenli, O. K.; Cakir, B., Ultrasonographic evaluation of thyroid nodules: comparison of ultrasonographic, cytological, and histopathological findings. *Endocrine* **2009,** *36* (3), 464-472.

112. Lan, L.; Luo, Y.; Zhou, M.; Huo, L.; Chen, H.; Zuo, Q.; Deng, W., Comparison of Diagnostic Accuracy of Thyroid Cancer With Ultrasound-Guided Fine-Needle Aspiration and Core-Needle Biopsy: A Systematic Review and Meta-Analysis. *Frontiers in Endocrinology* **2020,** *11* (44).

113. Renshaw, A. A.; Gould, E. W.; Russ, G.; Poller, D. N., Thyroid FNA: Is cytopathologist review of ultrasound features useful? *Cancer Cytopathology* **2020,** *128* (8), 523-527.

114. Lew, J. I.; Solorzano, C. C., Use of Ultrasound in the Management of Thyroid Cancer. *The Oncologist* **2010,** *15* (3), 253-258.

115. Wong, K. T.; Ahuja, A. T., Ultrasound of thyroid cancer. *Cancer Imaging* **2005,** *5* (1), 157-166.

116. Frates, M. C.; Benson, C. B.; Charboneau, J. W.; Cibas, E. S.; Clark, O. H.; Coleman, B. G.; Cronan, J. J.; Doubilet, P. M.; Evans, D. B.; Goellner, J. R.; Hay, I. D.; Hertzberg, B. S.; Intenzo, C. M.; Jeffrey, R. B.; Langer, J. E.; Larsen; Reed, P.; Mandel, S. J.; Middleton, W. D.; Reading, C. C.; Sherman, S. I.; Tessler, F. N., Management of Thyroid Nodules Detected at US Society of Radiologists in Ultrasound Consensus Conference Statement. *Ultrasound Quarterly* & **2006,** *22* (4), 231-238.

117. Yoo, W. S.; Choi, H. S.; Cho, S. W.; Moon, J. H.; Kim, K. W.; Park, H. J.; Park, S. Y.; Choi, S. I.; Choi, S. H.; Lim, S.; Yi, K. H.; Park, D. J.; Jang, H. C.; Park, Y. J., The role of ultrasound findings in the management of thyroid nodules with atypia or follicular lesions of undetermined significance. *Clinical Endocrinology* **2014,** *80* (5), 735-42.

118. Ginat, D. T.; Butani, D.; Giampoli, E. J.; Patel, N.; Dogra, V., Pearls and pitfalls of thyroid nodule sonography and fine-needle aspiration. *Ultrasound Quarterly* **2010,** *26* (3), 171-178.

119. Tuttle, R. M., Controversial Issues in Thyroid Cancer Management. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **2018,** *59* (8), 1187-1194.

120. Ozel, A.; Erturk, S. M.; Ercan, A.; Yilmaz, B.; Basak, T.; Karpat, Z.; Basak, M., The Diagnostic Efficiency of Ultrasound in Characterization of Thyroid Nodules: How Many Criteria Are Needed? *Medical ultrasonography* **2012,** *14* (1), 24-28.

121. Wang, Y.; Lei, K. R.; He, Y. P.; Li, X. L.; Ren, W. W.; Zhao, C. K.; Bo, X. W.; Wang, D.; Sun, C. Y.; Xu, H. X., Malignancy risk stratification of thyroid nodules: Comparisons of four ultrasound Thyroid Imaging Reporting and Data Systems in surgically resected nodules. *Scientific Reports* **2017,** *7* (1), 1-10.

122. Park, J.-Y.; Lee, H. J.; Jang, H. W.; Kim, H. K.; Yi, J. H.; Lee, W.; Kim, S. H., A Proposal for a Thyroid Imaging Reporting and Data System for Ultrasound Features of Thyroid Carcinoma. *Thyroid* **2009,** *19* (11), 1257-1264.

123. Lee, J. Y.; Na, D. G.; Yoon, S. J.; Gwon, H. Y.; Paik, W.; Kim, T.; Kim, J. Y., Ultrasound malignancy risk stratification of thyroid nodules based on the degree of hypoechogenicity and echotexture. *European Radiology* **2020,** *30* (3), 1653-1663.

124. Smith-Bindman, R.; Lebda, P.; Feldstein, V. A.; Sellami, D.; Goldstein, R. B.; Brasic, N.; Jin, C.; Kornak, J., Risk of thyroid cancer based on thyroid ultrasound imaging

characteristics: Results of a population-based study. *JAMA Internal Medicine* **2013,** *173* (19), 1788-1796.

125. Madani, G., Thyroid imaging. Arora, A.; Tolley, N. S.; Tuttle, M. R., Eds. Blackwell Publishing Limited: 2010; pp 36-43.

126. Papini, E.; Guglielmi, R.; Bianchini, A.; Crescenzi, A.; Taccogna, S.; Nardi, F.; Panunzi, C.; Rinaldi, R.; Toscano, V.; Pacella, C. M., Risk of Malignancy in Nonpalpable Thyroid Nodules: Predictive Value of Ultrasound and Color-Doppler Features. *The Journal of Clinical Endocrinology & Metabolism* **2002,** *87* (5), 1941-1946.

127. Baser, H.; Cakir, B.; Topaloglu, O.; Alkan, A.; Polat, S. B.; Dogan, H. T.; Yazicioğlu, M. O.; Aydin, C.; Ersoy, R., Diagnostic accuracy of Thyroid Imaging Reporting and Data System in the prediction of malignancy in nodules with atypia and follicular lesion of undetermined significance cytologies. *Clinical Endocrinology* **2017,** *86* (4), 584-590.

128. Kaliszewski, K.; Diakowska, D.; Wojtczak, B.; Forkasiewicz, Z., Evaluation of selected ultrasound features of thyroid nodules with atypia of undetermined significance/follicular lesion of undetermined significance for the Bethesda reporting system for thyroid cytology. *Cancer Manag Res* **2018,** *10*, 2223-2229.

129. Nabhan, F.; Ringel, M. D., Thyroid nodules and cancer management guidelines: Comparisons and controversies. 2017; Vol. 24, pp R13-R26.

130. Ginat, D. T.; Butani, D.; Giampoli, E. J.; Patel, N.; Dogra, V., Pearls and pitfalls of thyroid nodule sonography and fine-needle aspiration. 2010; Vol. 26, pp 171-178.

131. McHenry, C. R.; Phitayakorn, R., Follicular adenoma and carcinoma of the thyroid gland. *The oncologist* **2011,** *16* (5), 585-593.

132. Floridi, C.; Cellina, M.; Buccimazza, G.; Arrichiello, A.; Sacrini, A.; Arrigoni, F.; Pompili, G.; Barile, A.; Carrafiello, G., Ultrasound imaging classifications of thyroid nodules

for malignancy risk stratification and clinical management: state of the art. *Gland surgery* **2019,** *8* (Suppl 3), S233-S244.

133. Yoon, J. H.; Lee, H. S.; Kim, E.-K.; Moon, H. J.; Park, V. Y.; Kwak, J. Y., Pattern-based vs. score-based guidelines using ultrasound features have different strengths in risk stratification of thyroid nodules. *European Radiology* **2020,** *30* (7), 3793-3802.

134. Tessler, F. N.; Middleton, W. D.; Grant, E. G., ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS Committee. *J Am Coll Radiol* **2017,** *14*, 587.

135. Russ, G.; Bonnema, S. J.; Erdogan, M. F.; Durante, C.; Ngu, R.; Leenhardt, L., European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS. *Eur Thyroid J* **2017,** *6* (5), 225-237.

136. Shin, J. H.; Baek, J. H.; Chung, J.; Ha, E. J.; Kim, J. H.; Lee, Y. H.; Lim, H. K.; Moon, W. J.; Na, D. G.; Park, J. S.; Choi, Y. J.; Hahn, S. Y.; Jeon, S. J.; Jung, S. L.; Kim, D. W.; Kim, E. K.; Kwak, J. Y.; Lee, C. Y.; Lee, H. J.; Lee, J. H.; Lee, J. H.; Lee, K. H.; Park, S. W.; Sung, J. Y., Ultrasonography diagnosis and imaging-based management of thyroid nodules: Revised Korean society of thyroid radiology consensus statement and recommendations. *Korean Journal of Radiology* **2016,** *17* (3), 370-395.

137. Kim, P. H.; Suh, C. H.; Baek, J. H.; Chung, S. R.; Choi, Y. J.; Lee, J. H., Unnecessary thyroid nodule biopsy rates under four ultrasound risk stratification systems: a systematic review and meta-analysis. *European Radiology* **2021,** *31* (5), 2877-2885.

138. Grani, G.; Lamartina, L.; Ascoli, V., Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the “right” TIRADS. *J Clin Endocrinol Metab* **2019,** *104*, 95.

139. Chen, H.; Ye, J.; Song, J.; You, Y.; Chen, W.; Liu, Y., Comparison of Different Ultrasound Classification Systems of Thyroid Nodules for Identifying Malignant Potential: A Cross-sectional Study. *Clinics (Sao Paulo)* **2021,** *76*, e2126-e2126.

140. Castellana, M.; Castellana, C.; Treglia, G.; Giorgino, F.; Giovanella, L.; Russ, G.; Trimboli, P., Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. A meta-analysis. *The Journal of Clinical Endocrinology & Metabolism* **2020,** *105* (5).

141. Schenke, S.; Klett, R.; Seifert, P.; Kreissl, C. M.; Görges, R.; Zimny, M., Diagnostic Performance of Different Thyroid Imaging Reporting and Data Systems (Kwak-TIRADS, EU-TIRADS and ACR TI-RADS) for Risk Stratification of Small Thyroid Nodules (≤10 mm). *Journal of Clinical Medicine* **2020,** *9* (1).

142. Kwak, J. Y.; Jung, I.; Baek, J. H.; Baek, S. M.; Choi, N.; Choi, Y. J.; Jung, S. L.; Kim, E. K.; Kim, J. A.; Kim, J. H.; Kim, K. S.; Lee, J. H.; Lee, J. H.; Moon, H. J.; Moon, W. J.; Park, J. S.; Ryu, J. H.; Shin, J. H.; Son, E. J.; Sung, J. Y.; Na, D. G., Image reporting and characterization system for ultrasound features of thyroid nodules: Multicentric Korean retrospective study. *Korean Journal of Radiology* **2013,** *14* (1), 110-117.

143. Migda, B.; Migda, M.; Migda, M. S., A systematic review and meta-analysis of the Kwak TIRADS for the diagnostic assessment of indeterminate thyroid nodules. *Clinical Radiology* **2019,** *74* (2), 123-130.

144. Ha, E. J.; Na, D. G.; Baek, J. H.; Sung, J. Y.; Kim, J. H.; Kang, S. Y., US fine-needle aspiration biopsy for thyroid malignancy: diagnostic performance of seven society guidelines applied to 2000 thyroid nodules. *Radiology* **2018,** *287*, 893.

145. Shreyamsa, M.; Mishra, A.; Ramakant, P.; Parihar, A.; Singh, K. R.; Rana, C.; Mouli, S., Comparison of Multimodal Ultrasound Imaging with Conventional Ultrasound Risk

Stratification Systems in Presurgical Risk Stratification of Thyroid Nodules. *Indian J Endocrinol Metab* **2020,** *24* (6), 537-542.

146. Grani, G.; Lamartina, L.; Ascoli, V.; Bosco, D.; Biffoni, M.; Giacomelli, L.; Maranghi, M.; Falcone, R.; Ramundo, V.; Cantisani, V.; Filetti, S.; Durante, C., Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: Toward the "Right" TIRADS. *Journal of Clinical Endocrinology and Metabolism* **2019,** *104* (1), 95-102.

147. Grani, G.; Lamartina, L.; Ascoli, V.; Bosco, D.; Nardi, F.; D'Ambrosio, F.; Rubini, A.; Giacomelli, L.; Biffoni, M.; Filetti, S.; Durante, C.; Cantisani, V., Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. *Endocrine* **2017,** *57* (2), 256-261.

148. Giuliano, S.; Mirabelli, M.; Chiefari, E.; Vergine, M.; Gervasi, R.; Brunetti, F. S.; Innaro, N.; Donato, G.; Aversa, A.; Brunetti, A., Malignancy Analyses of Thyroid Nodules in Patients Subjected to Surgery with Cytological- and Ultrasound-Based Risk Stratification Systems. *Endocrines* **2020,** *1* (2), 102-118.

149. Slowinska-Klencka, D.; Wysocka-Konieczna, K.; Klencki, M.; Popowicz, B., Diagnostic Value of Six Thyroid Imaging Reporting and Data Systems (TIRADS) in Cytologically Equivocal Thyroid Nodules. *J Clin Med* **2020,** *9* (7), 2281.

150. Trimboli, P.; Fulciniti, F.; Zilioli, V.; Ceriani, L.; Giovanella, L., Accuracy of international ultrasound risk stratification systems in thyroid lesions cytologically classified as indeterminate. *Diagnostic Cytopathology* **2017,** *45* (2), 113-117.

151. Moon, H. J.; Kim, E.-K.; Yoon, J. H.; Kwak, J. Y., Malignancy Risk Stratification in Thyroid Nodules with Nondiagnostic Results at Cytologic Examination: Combination of Thyroid Imaging Reporting and Data System and the Bethesda System. *Radiology* **2015,** *274* (1), 287-295.

152. Ringel, M. D., Thyroid Cancer. 2016; pp 991-999.

153. Holden, A., The role of colour and duplex Doppler ultrasound in the assessment of thyroid nodules. *Australasian Radiology* **1995,** *39* (4), 343-349.

154. Papini, E.; Pacella, C. M.; Frasoldati, A.; Hegedüs, L.; Papini, E.; Pacella, C. M.; Frasoldati, A.; Hegedüs, L., Ultrasonic Imaging of the Thyroid Gland. Wartofsky, L.; Van Nostrand, D., Eds. Springer Science: New York, NY, 2016; pp 293-313.

155. Özel, D.; Özel, B. D., Evaluation of diagnostic value of conventional and color doppler ultrasound with elastography strain ratios in differentiation between benign and malignant lymph nodes. *Polish Journal of Radiology* **2018,** *83*, 32-36.

156. Ebeed, A. E.; Romeih, M. A. E.; Refat, M. M.; Salah, N. M., Role of ultrasound, color doppler, elastography and micropure imaging in differentiation between benign and malignant thyroid nodules. *Egyptian Journal of Radiology and Nuclear Medicine* **2017,** *48* (3), 603-610.

157. Frates, M. C.; Benson, C. B.; Doubilet, P. M.; Cibas, E. S.; Marqusee, E., Can color doppler sonography aid in the prediction of malignancy of thyroid nodules? *Journal of Ultrasound in Medicine* **2003,** *22* (2), 127-131.

158. Darvish, L.; Khezri, M.; Teshnizi, S. H.; Roozbeh, N.; Dehkordi, J. G.; Amraee, A., Color Doppler ultrasonography diagnostic value in detection of malignant nodules in cysts with pathologically proven thyroid malignancy: a systematic review and meta-analysis. *Clinical and Translational Oncology* **2019,** *21* (12), 1712-1729.

159. Rago, T.; Vitti, P.; Chiovato, L.; Mazzeo, S.; De Liperi, A.; Miccoli, P.; Viacava, P.; Bogazzi, F.; Martino, E.; Pinchera, A., Role of conventional ultrasonography and color flow-doppler sonography in predicting malignancy in 'cold' thyroid nodules. *European Journal of Endocrinology* **1998,** *138* (1), 41-46.

160. Hong, Y.-r.; Wu, Y.-l.; Luo, Z.-y.; Wu, N.-b.; Liu, X.-m., Impact of nodular size on the predictive values of gray-scale, color-Doppler ultrasound, and sonoelastography for

assessment of thyroid nodules. *Journal of Zhejiang University. Science. B* **2012,** *13* (9), 707-16.

161. Chammas, M. C.; Gerhard, R.; Oliveira, I. R. S. D.; Widman, A.; Barros, N. D.; Durazzo, M.; Ferraz, A.; Cerri, G. G., Thyroid nodules: Evaluation with power Doppler and duplex Doppler ultrasound. *Otolaryngology–Head and Neck Surgery* **2005,** *132* (6), 874-882.

162. Slapa, R. Z.; Jakubowski, W. S.; Slowinska-Srzednicka, J.; Szopinski, K. T., Advantages and disadvantages of 3D ultrasound of thyroid nodules including thin slice volume rendering. *Thyroid research* **2011,** *4* (1), 1-1.

163. Milas, M.; Mandel, S. J.; Langer, J. E., *Advanced Thyroid and Parathyroid Ultrasound*. 2017.

164. Moon, H. J.; Kwak, J. Y.; Kim, M. J.; Son, E. J.; Kim, E.-K., Can Vascularity at Power Doppler US Help Predict Thyroid Malignancy? 1. **2010,** *255* (1), 260-269.

165. Rosario, P. W.; da Silva, A. L.; Borges, M. A. R.; Calsolari, M. R., Is Doppler ultrasound of additional value to gray-scale ultrasound in differentiating malignant and benign thyroid nodules? *Arch. Endocrinol. Metab.* **2015,** *59* (1), 79-83.

166. Chung, J.; Lee, Y. J.; Choi, Y. J.; Ha, E. J.; Suh, C. H.; Choi, M.; Baek, J. H.; Na, D. G., Clinical applications of Doppler ultrasonography for thyroid disease: consensus statement by the Korean Society of Thyroid Radiology. *Ultrasonography* **2020,** *39* (4), 315-330.

167. Tamsel, S.; Demirpolat, G.; Erdogan, M.; Nart, D.; Karadeniz, M.; Uluer, H.; Ozgen, A. G., Power Doppler US patterns of vascularity and spectral Doppler US parameters in predicting malignancy in thyroid nodules. *Clinical Radiology* **2007,** *62* (3), 245-251.

168. Ying, M.; Cheng, S. C. H.; Ahuja, A. T., Diagnostic Accuracy of Computer-Aided Assessment of Intranodal Vascularity in Distinguishing Different Causes of Cervical Lymphadenopathy. *Ultrasound in Medicine and Biology* **2016,** *42* (8), 2010-2016.

169. Lam, J.; Ying, M.; Cheung, S. Y.; Yeung, K. H.; Yu, P. H.; Cheng, H. C.; Ahuja, A. T., A Comparison of the Diagnostic Accuracy and Reliability of Subjective Grading and Computer-Aided Assessment of Intranodal Vascularity in Differentiating Metastatic and Reactive Cervical Lymphadenopathy. *Ultraschall in der Medizin* **2016,** *37* (1), 63-67.

170. Toomatari, S. B. M.; Mohammadi, A.; Sepehrvand, N.; Toomatari, S. E. M.; Ghasemi-Rad, M.; Shamspour, S. Z.; Rezayi, S.; Toubaei, M.; Sarabi, Z. K., A novel computerised quantification of thyroid vascularity in the differentiation of malignant and benign thyroid nodules. *Polish Journal of Radiology* **2019,** *84*, E517-E521.

171. Yoon, J. H.; Shin, H. J.; Kim, E.-K.; Moon, H. J.; Roh, Y. H.; Kwak, J. Y., Quantitative Evaluation of Vascularity Using 2-D Power Doppler Ultrasonography May Not Identify Malignancy of the Thyroid. *Ultrasound in Medicine & Biology* **2015,** *41* (11), 2873-2883.

172. MacHado, P.; Segal, S.; Lyshchik, A.; Forsberg, F., A novel microvascular flow technique: Initial results in thyroids. *Ultrasound Quarterly* **2016,** *32* (1), 67-74.

173. Xiao, X. Y.; Chen, X.; Guan, X. F.; Wu, H.; Qin, W.; Luo, B. M., Superb microvascular imaging in diagnosis of breast lesions: A comparative study with contrast-enhanced ultrasonographic microvascular imaging. *British Journal of Radiology* **2016,** *89* (1066).

174. Park, A. Y.; Seo, B. K., Up-to-date Doppler techniques for breast tumor vascularity: superb microvascular imaging and contrast-enhanced ultrasound. **2018,** *37* (2), 98-106.

175. Lu, R.; Meng, Y.; Zhang, Y.; Zhao, W.; Wang, X.; Jin, M.; Guo, R., Superb microvascular imaging (SMI) compared with conventional ultrasound for evaluating thyroid nodules. *BMC Medical Imaging* **2017,** *17* (1), 65-65.

176. Cappelli, C.; Pirola, I.; Gandossi, E.; Marini, F.; Cristiano, A.; Casella, C.; Lombardi, D.; Agosti, B.; Ferlin, A.; Castellano, M., Ultrasound microvascular blood flow

evaluation: A new tool for the management of thyroid nodule? *International Journal of Endocrinology* **2019,** *2019*, 7874890.

177. Kong, J.; Li, J.-c.; Wang, H.-y.; Wang, Y.-h.; Zhao, R.-n.; Zhang, Y.; Jin, J., Role of Superb Micro-Vascular Imaging in the Preoperative Evaluation of Thyroid Nodules: Comparison With Power Doppler Flow Imaging. *Journal of Ultrasound in Medicine* **2017,** *36* (7), 1329-1337.

178. Pei, S.; Cong, S.; Zhang, B.; Liang, C.; Zhang, L.; Liu, J.; Guo, Y.; Zhang, S., Diagnostic value of multimodal ultrasound imaging in differentiating benign and malignant TI-RADS category 4 nodules. *Int. J. Clin. Oncol.* **2019,** *24* (6), 632-639.

179. Yoon, J. H.; Kim, E.-K.; Kwak, J. Y.; Park, V. Y.; Moon, H. J., Application of Various Additional Imaging Techniques for Thyroid Ultrasound: Direct Comparison of Combined Various Elastography and Doppler Parameters to Gray-Scale Ultrasound in Differential Diagnosis of Thyroid Nodules. *Ultrasound in Medicine & Biology* **2018,** *44* (8), 1679-1686.

180. Chng, C. L.; Tan, H. C.; Too, C. W.; Lim, W. Y.; Chiam, P. P. S.; Zhu, L.; Nadkarni, N. V.; Lim, A. Y. Y., Diagnostic performance of ATA, BTA and TIRADS sonographic patterns in the prediction of malignancy in histologically proven thyroid nodules. *Singapore medical journal* **2018,** *59* (11), 578-583.

181. Sigrist, R. M. S.; Liau, J.; Kaffas, A. E.; Chammas, M. C.; Willmann, J. K.; Willmann, J. K., Ultrasound Elastography: Review of Techniques and Clinical Applications. *Theranostics* **2017,** *7* (5), 1303-1329.

182. Shiina, T.; Nightingale, K. R.; Palmeri, M. L.; Hall, T. J.; Bamber, J. C.; Barr, R. G.; Castera, L.; Choi, B. I.; Chou, Y. H.; Cosgrove, D.; Dietrich, C. F.; Ding, H.; Amy, D.; Farrokh, A.; Ferraioli, G.; Filice, C.; Friedrich-Rust, M.; Nakashima, K.; Schafer, F.; Sporea, I.; Suzuki, S.; Wilson, S.; Kudo, M., WFUMB guidelines and recommendations for

clinical use of ultrasound elastography: Part 1: Basic principles and terminology. *Ultrasound in Medicine and Biology* **2015,** *41* (5), 1126-1147.

183. Cosgrove, D.; Barr, R.; Bojunga, J.; Cantisani, V.; Chammas, M. C.; Dighe, M.; Vinayak, S.; Xu, J. M.; Dietrich, C. F., WFUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography: Part 4. Thyroid. 2017; Vol. 43, pp 4-26.

184. Zhang, B.; Ma, X.; Wu, N.; Liu, L.; Liu, X.; Zhang, J.; Yang, J.; Niu, T., Shear Wave Elastography for Differentiation of Benign and Malignant Thyroid Nodules. *Journal of Ultrasound in Medicine* **2013,** *32* (12), 2163-2169.

185. Samir, A. E.; Dhyani, M.; Anvari, A.; Prescott, J.; Halpern, E. F.; Faquin, W. C.; Stephen, A., Shear-Wave Elastography for the Preoperative Risk Stratification of Follicular-patterned Lesions of the Thyroid: Diagnostic Accuracy and Optimal Measurement Plane. *Radiology* **2015,** *277* (2), 565-573.

186. Tian, W.; Hao, S.; Gao, B.; Jiang, Y.; Zhang, S.; Guo, L.; Luo, D., Comparison of Diagnostic Accuracy of Real-Time Elastography and Shear Wave Elastography in Differentiation Malignant From Benign Thyroid Nodules. *Medicine* **2015,** *94* (52), e2312-e2312.

187. Sohail, S.; Uz Zaman, S. K., Diagnostic value of mean elasticity index as a quantitative shear wave elastography parameter for prediction of malignancy in small suspicious solid thyroid nodules. *Journal of the College of Physicians and Surgeons Pakistan* **2020,** *30* (7), 683-687.

188. Kim, H. J.; Kwak, M. K.; Choi, I. H.; Jin, S.-Y.; Park, H. K.; Byun, D. W.; Suh, K.; Yoo, M. H., Utility of shear wave elastography to detect papillary thyroid carcinoma in thyroid nodules: efficacy of the standard deviation elasticity. *The Korean journal of internal medicine* **2019,** *34* (4), 850-857.

189. Gregory, A.; Bayat, M.; Kumar, V.; Denis, M.; Kim, B. H.; Webb, J.; Meixner, D. D.; Ryder, M.; Knudsen, J. M.; Chen, S.; Fatemi, M.; Alizad, A., Differentiation of Benign and Malignant Thyroid Nodules by Using Comb-push Ultrasound Shear Elastography: A Preliminary Two-plane View Study. *Academic Radiology* **2018,** *25* (11), 1388-1397.

190. Baig, F.; Liu, S.; Lam, H.-C.; Yip, S.-P.; Law, H.; Ying, M., Shear Wave Elastography Combining with Conventional Grey Scale Ultrasound Improves the Diagnostic Accuracy in Differentiating Benign and Malignant Thyroid Nodules. *Applied Sciences* **2017,** *7* (11), 1103-1103.

191. Filho, R. H. C.; Pereira, F. L.; Iared, W., Diagnostic Accuracy Evaluation of Two-Dimensional Shear Wave Elastography in the Differentiation Between Benign and Malignant Thyroid Nodules: Systematic Review and Meta-analysis. *J Ultrasound Med* **2020,** *39* (9), 1729-1741.

192. Cantisani, V.; Ulisse, S.; Guaitoli, E.; de Vito, C.; Caruso, R.; Mocini, R.; D'Andrea, V.; Ascoli, V.; Antonaci, A.; Catalano, C.; Nardi, F.; Redler, A.; Ricci, P.; de Antoni, E.; Sorrenti, S., Q-Elastography in the Presurgical Diagnosis of Thyroid Nodules with Indeterminate Cytology. *PLoS ONE* **2012,** *7* (11), e50725.

193. Rago, T.; Scutari, M.; Santini, F.; Loiacono, V.; Piaggi, P.; Di Coscio, G.; Basolo, F.; Berti, P.; Pinchera, A.; Vitti, P., Real-time elastosonography: useful tool for refining the presurgical diagnosis in thyroid nodules with indeterminate or nondiagnostic cytology. *J Clin Endocrinol Metab* **2010,** *95* (12), 5274-80.

194. Lippolis, P. V.; Tognini, S.; Materazzi, G.; Polini, A.; Mancini, R.; Ambrosini, C. E.; Dardano, A.; Basolo, F.; Seccia, M.; Miccoli, P.; Monzani, F., Is elastography actually useful in the presurgical selection of thyroid nodules with indeterminate cytology? *Journal of Clinical Endocrinology and Metabolism* **2011,** *96* (11), E1826-30.

195. Trimboli, P.; Treglia, G.; Sadeghi, R.; Romanelli, F.; Giovanella, L., Reliability of real-time elastography to diagnose thyroid nodules previously read at FNAC as indeterminate: a meta-analysis. *Endocrine* **2015,** *50* (2), 335-343.

196. Qiu, Y.; Xing, Z.; Liu, J.; Peng, Y.; Zhu, J.; Su, A., Diagnostic reliability of elastography in thyroid nodules reported as indeterminate at prior fine-needle aspiration cytology (FNAC): a systematic review and Bayesian meta-analysis. *Eur Radiol* **2020,** *30* (12), 6624-6634.

197. Zhang, L.; Gu, J.; Zhao, Y.; Zhu, M.; Wei, J.; Zhang, B., The role of multimodal ultrasonic flow imaging in Thyroid Imaging Reporting and Data System (TI-RADS) 4 nodules. *Gland Surg* **2020,** *9* (5), 1469-1477.

198. Stoian, D.; Ivan, V.; Sporea, I.; Florian, V.; Mozos, I.; Navolan, D.; Nemescu, D., Advanced Ultrasound Application - Impact on Presurgical Risk Stratification of the Thyroid Nodules. *Ther Clin Risk Manag* **2020,** *16*, 21-30.

199. Yoon, J. H.; Kim, E. K.; Kwak, J. Y.; Park, V. Y.; Moon, H. J., Application of Various Additional Imaging Techniques for Thyroid Ultrasound: Direct Comparison of Combined Various Elastography and Doppler Parameters to Gray-Scale Ultrasound in Differential Diagnosis of Thyroid Nodules. *Ultrasound Med Biol* **2018,** *44* (8), 1679-1686.

200. Giusti, M.; Campomenosi, C.; Gay, S.; Massa, B.; Silvestri, E.; Monti, E.; Turtulici, G., The use of semi-quantitative ultrasound elastosonography in combination with conventional ultrasonography and contrast-enhanced ultrasonography in the assessment of malignancy risk of thyroid nodules with indeterminate cytology. *Thyroid Research* **2014,** *7* (1), 9.

201. Borlea, A.; Stoian, D.; Cotoi, L.; Sporea, I.; Lazar, F.; Mozos, I., Thyroid Multimodal Ultrasound Evaluation—Impact on Presurgical Diagnosis of Intermediate Cytology Cases. *Applied Sciences* **2020,** *10* (10), 3439.

202. Wu, J. X.; Livhits, M.; Sepahdari, A.; Yeh, M. W., Cross-Sectional Imaging for the Evaluation of Thyroid Nodules and Cancer. Roman, S. A.; Sosa, J. A.; Solorzano, C. C., Eds. Springer International Publishing Switzerland: Los Angeles, CA, 2017; pp 93-102.

203. Roy, M.; Chen, H.; Sippel, R. S., Current understanding and management of medullary thyroid cancer. *The oncologist* **2013,** *18* (10), 1093-100.

204. Miyakoshi, A.; Dalley, R. W.; Anzai, Y., Magnetic resonance imaging of thyroid cancer. *Topics in Magnetic Resonance Imaging* **2007,** *18* (4), 293-302.

205. Henrichsen, T. L., Imaging for Thyroid Nodules. Gharib, H., Ed. Springer International Publishing AG: Rochester, MN, 2018; pp 63-79.

206. Jelinek, J.; Young, R.; Smith Iii, L. O.; Burman, K. D., MR and CT Imaging of Thyroid Cancer. Wartofsky, L.; Van Nostrand, D., Eds. Springer+Business Media: New York, 2016; pp 515-524.

207. Shetty, S. K.; Maher, M. M.; Hahn, P. F.; Halpern, E. F.; Aquino, S. L.; Sk, S.; Mm, M.; Pf, H.; Ef, H.; Sl, A., Significance of Incidental Thyroid Lesions Detected on CT: Correlation Among CT, Sonography, and Pathology. *AJR* **2006,** *187* (5), 1349-1356.

208. Esposito, G., Radionuclide Imaging of Medullary Carcinoma. Wartofsky, L.; Van Nostrand, D., Eds. Spinger Science+Business Media: New York, 2016; pp 873-880.

209. Sharma, B.; Martin, A.; Stanway, S.; Johnston, S. R. D.; Constantinidou, A., Imaging in oncology—over a century of advances. *Nature Reviews Clinical Oncology* **2012,** *9* (12), 728-737.

210. Chaudhary, V.; Bano, S., Imaging of the thyroid: Recent advances. *Indian Journal of Endocrinology and Metabolism* **2012,** *16* (3), 371-371.

211. Hoang, J. K.; Branstetter Iv, B. F.; Gafton, A. R.; Lee, W. K.; Glastonbury, C. M., Imaging of thyroid carcinoma with CT and MRI: Approaches to common scenarios. *Cancer Imaging* **2013,** *13* (1), 128-139.

212. Zimmer, L. A.; McCook, B.; Meltzer, C.; Fukui, M.; Bascom, D.; Snyderman, C.; Townsend, D. W.; Johnson, J. T., Combined positron emission tomography/computed tomography imaging of recurrent thyroid cancer. *Otolaryngology - Head and Neck Surgery* **2003,** *128* (2), 178-184.

213. Durr, E. S.; Rahim, K., Thyroid Nodule Imaging, Status and Limitations. *Asia Oceania journal of nuclear medicine & biology* **2015,** *3* (1), 50-7.

214. Nostrand, D. V.; Schneider, M.; Acio, E. R., Radionuclide Imaging of Thyroid Nodules. Wartofsky, L.; Van Nostrand, D., Eds. Springer+Business Media: New York, 2016; pp 315-322.

215. Wong, K. T.; Choi, F. P. T.; Lee, Y. Y. P.; Ahuja, A. T., Current role of radionuclide imaging in differentiated thyroid cancer. *Cancer Imaging* **2008,** *8*, 159-162.

216. Iagaru, A.; McDougall, I. R., Positron Emission Tomography- Computed Tomography (PET-CT and PET) in Well-Differentiated Thyroid Cancer. Wartofsky, L.; Van Nostrand, D., Eds. Spinger Science+Business Media: New York, 2016; pp 487-504.

217. Zoller, M.; Kohlfuerst, S.; Igerc, I.; Kresnik, E.; Gallowitsch, H.-J.; Gomez, I.; Lind, P., Combined PET/CT in the follow-up of differentiated thyroid carcinoma: what is the impact of each modality? *European Journal of Nuclear Medicine and Molecular Imaging* **2007,** *34* (4), 487-495.

218. Garcia, C. A., 18F Fluorodeoxyglucose PET and Thyroid Nodules. Wartofsky, L.; Van Nostrand, D., Eds. Spinger Science+Business Media: New York, 2016; pp 323-330.

219. Scappaticcio, L.; Piccardo, A.; Treglia, G.; Poller, D. N.; Trimboli, P., The dilemma of 18F-FDG PET/CT thyroid incidentaloma: what we should expect from FNA. A systematic review and meta-analysis. *Endocrine* **2021,** *73* (3), 540-549.

220. Piccardo, A.; Puntoni, M.; Dezzana, M.; Bottoni, G.; Foppiani, L.; Marugo, A.; Catrambone, U.; Ugolini, M.; Sola, S.; Gatto, M.; Treglia, G.; Giovanella, L.; Trimboli, P.,

Indeterminate thyroid nodules. The role of 18F-FDG PET/CT in the "era" of ultrasonography risk stratification systems and new thyroid cytology classifications. *Endocrine* **2020,** *69* (3), 553-561.

221. Ho, T.-Y.; Liou, M.-J.; Lin, K.-J.; Yen, T.-C., Prevalence and significance of thyroid uptake detected by 18F-FDG PET. *Endocrine* **2011,** *40* (2), 297-302.

222. Palmedo, H.; Bucerius, J.; Joe, A.; Strunk, H.; Hortling, N.; Meyka, S.; Roedel, R.; Wolff, M.; Wardelmann, E.; Biersack, H.-J.; Jaeger, U., Integrated PET/CT in differentiated thyroid cancer: diagnostic accuracy and impact on patient management. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **2006,** *47* (4), 616-24.

223. A, I.; J.E, K.; I.R, M.; Iagaru, A.; Kalinyak, J. E.; McDougall, I. R., F-18 FDG PET/CT in the management of thyroid cancer. *Clinical nuclear medicine* **2007,** *32* (9), 690-695.

224. Piccardo, A.; Trimboli, P.; Foppiani, L.; Treglia, G.; Ferrarazzo, G.; Massollo, M.; Bottoni, G.; Giovanella, L., PET/CT in thyroid nodule and differentiated thyroid cancer patients. The evidence-based state of the art. *Reviews in Endocrine and Metabolic Disorders* **2019,** *20* (1), 47-64.

225. Mosci, C.; Iagaru, A., PET/CT imaging of thyroid cancer. *Clinical Nuclear Medicine* **2011,** *36* (12), e180-e185.

226. Choi, J. Y.; Lee, K. S.; Kim, H.-J.; Shim, Y. M.; Kwon, O. J.; Park, K.; Baek, C.-H.; Chung, J. H.; Lee, K.-H.; Kim, B.-T., Focal thyroid lesions incidentally identified by integrated 18F-FDG PET/CT: clinical significance and improved characterization. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **2006,** *47* (4), 609-15.

227. Demir, O.; Kose, N.; Ozkan, E.; Unluturk, U.; Aras, G.; Erdogan, M. F., Clinical significance of thyroid incidentalomas identified by 18F-FDG PET/CT: Correlation of

ultrasonograpy findings with cytology results. *Nuclear Medicine Communications* **2016,** *37* (7), 715-720.

228. Chang, T.-C., The Role of Computer-aided Detection and Diagnosis System in the Differential Diagnosis of Thyroid Lesions in Ultrasonography. *Journal of Medical Ultrasound* **2015,** *23* (4), 177-184.

229. Faust, O.; Acharya, U. R.; Tamura, T., Formal Design Methods for Reliable Computer-Aided Diagnosis: A Review. *IEEE Reviews in Biomedical Engineering* **2012,** *5*, 15-28.

230. Acharya, U. R.; Swapna, G.; Sree, S. V.; Molinari, F.; Gupta, S.; Bardales, R. H.; Witkowska, A.; Suri, J. S., A Review on Ultrasound-Based Thyroid Cancer Tissue Characterization and Automated Classification. *Technology in Cancer Research & Treatment* **2014,** *13* (4), 289-301.

231. Ying, M.; Ng, D. K.; Yung, D. M.; Lee, E. S., A semi-quantitative approach to compare high-sensitivity power Doppler sonography and conventional power Doppler sonography in the assessment of thyroid vascularity. *Thyroid* **2009,** *19* (11), 1265-1269.

232. Faust, O.; Acharya, U. R.; Meiburger, K. M.; Molinari, F.; Koh, J. E. W.; Yeong, C. H.; Kongmebhol, P.; Ng, K. H., Comparative assessment of texture features for the identification of cancer in ultrasound images: a review. 2018; Vol. 38, pp 275-296.

233. Ha, E. J.; Baek, J. H., Applications of machine learning and deep learning to thyroid imaging: where do we stand? *Ultrasonography* **2021,** *40* (1), 23-29.

234. Zhao, W.-j.; Fu, L.-R.; Huang, Z.-M.; Zhu, J.-Q.; Ma, B.-Y., Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound. *Medicine* **2019,** *98* (32), 1-13.

235. Wang, L.; Yang, S.; Yang, S.; Zhao, C.; Tian, G.; Gao, Y.; Chen, Y.; Lu, Y., Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. *World Journal of Surgical Oncology* **2019,** *17* (1), 1-9.

236. Jin, Z.; Zhu, Y.; Zhang, S.; Xie, F.; Zhang, M.; Zhang, Y.; Tian, X.; Zhang, J.; Luo, Y.; Cao, J., Ultrasound Computer-Aided Diagnosis (CAD) Based on the Thyroid Imaging Reporting and Data System (TI-RADS) to Distinguish Benign from Malignant Thyroid Nodules and the Diagnostic Performance of Radiologists with Different Diagnostic Experience. *Med Sci Monit* **2020,** *26*, e918452.

237. Xu, L.; Gao, J.; Wang, Q.; Yin, J.; Yu, P.; Bai, B.; Pei, R.; Chen, D.; Yang, G.; Wang, S.; Wan, M., Computer-Aided Diagnosis Systems in Diagnosing Malignant Thyroid Nodules on Ultrasonography: A Systematic Review and Meta-Analysis. *European Thyroid Journal* **2020,** *9* (4), 186-193.

238. Chambara, N.; Ying, M., The Diagnostic Efficiency of Ultrasound Computer-Aided Diagnosis in Differentiating Thyroid Nodules: A Systematic Review and Narrative Synthesis. *Cancers (Basel)* **2019,** *11* (11).

239. Choi, Y. J.; Baek, J. H.; Park, H. S.; Shim, W. H.; Kim, T. Y.; Shong, Y. K.; Lee, J. H., A Computer-Aided Diagnosis System Using Artificial Intelligence for the Diagnosis and Characterization of Thyroid Nodules on Ultrasound: Initial Clinical Assessment. *Thyroid* **2017,** *27* (4), 546-552.

240. Gao, L.; Liu, R.; Jiang, Y.; Song, W.; Wang, Y.; Liu, J.; Wang, J.; Wu, D.; Li, S.; Hao, A.; Zhang, B., Computer-aided system for diagnosing thyroid nodules on ultrasound: A comparison with radiologist-based clinical assessments. *Head and Neck* **2018,** *40* (4), 778-783.

241. Yoo, Y. J.; Ha, E. J.; Cho, Y. J.; Kim, H. L.; Han, M.; Kang, S. Y., Computer-Aided Diagnosis of Thyroid Nodules via Ultrasonography: Initial Clinical Experience. *Korean Journal of Radiology* **2018,** *19* (4), 665-672.

242. Reverter, J. L.; Vázquez, F.; Puig-Domingo, M., Diagnostic Performance Evaluation of a Computer-Assisted Imaging Analysis System for Ultrasound Risk Stratification of Thyroid Nodules. *American Journal of Roentgenology* **2019**, 1-6.

243. Xia, S.; Yao, J.; Zhou, W.; Dong, Y.; Xu, S.; Zhou, J.; Zhan, W., A computer-aided diagnosing system in the evaluation of thyroid nodules—experience in a specialized thyroid center. *World Journal of Surgical Oncology* **2019,** *17* (1), 210.

244. Kim, H. L.; Ha, E. J.; Han, M., Real-World Performance of Computer-Aided Diagnosis System for Thyroid Nodules Using Ultrasonography. *Ultrasound Med Biol* **2019,** *45* (10), 2672-2678.

245. Han, M.; Ha, E. J.; Park, J. H., Computer-Aided Diagnostic System for Thyroid Nodules on Ultrasonography: Diagnostic Performance Based on the Thyroid Imaging Reporting and Data System Classification and Dichotomous Outcomes. *AJNR Am J Neuroradiol* **2021,** *42* (3), 559-565.

246. Ye, F. Y.; Lyu, G. R.; Li, S. Q.; You, J. H.; Wang, K. J.; Cai, M. L.; Su, Q. C., Diagnostic Performance of Ultrasound Computer-Aided Diagnosis Software Compared with That of Radiologists with Different Levels of Expertise for Thyroid Malignancy: A Multicenter Prospective Study. *Ultrasound Med Biol* **2021,** *47* (1), 114-124.

247. Gitto, S.; Grassi, G.; De Angelis, C.; Monaco, C. G.; Sdao, S.; Sardanelli, F.; Sconfienza, L. M.; Mauri, G., A computer-aided diagnosis system for the assessment and characterization of low-to-high suspicion thyroid nodules on ultrasound. *Radiologia Medica* **2019,** *124* (2), 118-125.

248. Li, T.; Jiang, Z.; Lu, M.; Zou, S.; Wu, M.; Wei, T.; Wang, L.; Li, J.; Hu, Z.; Cheng, X.; Liao, J., Computer-aided diagnosis system of thyroid nodules ultrasonography: Diagnostic performance difference between computer-aided diagnosis and 111 radiologists. *Medicine* **2020,** *99* (23), e20634-e20634.

249. Fresilli, D.; Grani, G.; De Pascali, M. L.; Alagna, G.; Tassone, E.; Ramundo, V.; Ascoli, V.; Bosco, D.; Biffoni, M.; Bononi, M.; D'Andrea, V.; Frattaroli, F.; Giacomelli, L.; Solskaya, Y.; Polti, G.; Pacini, P.; Guiban, O.; Gallo Curcio, R.; Caratozzolo, M.;

Cantisani, V., Computer-aided diagnostic system for thyroid nodule sonographic evaluation outperforms the specificity of less experienced examiners. *J Ultrasound* **2020,** *23* (2), 169-174.

250. Li, T. T.; Jiang, Z. R.; Lu, M.; Zou, S. B.; Wu, M. G.; Wei, T.; Wang, L.; Li, J.; Hu, Z. Y.; Cheng, X. Q.; Liao, J. F., Computer-aided diagnosis system of thyroid nodules ultrasonography Diagnostic performance difference between computer-aided diagnosis and 111 radiologists. *Medicine* **2020,** *99* (23).

251. Jeong, E. Y.; Kim, H. L.; Ha, E. J.; Park, S. Y.; Cho, Y. J.; Han, M., Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. *Eur Radiol* **2019,** *29* (4), 1978-1985.

252. Wu, M. H.; Chen, K. Y.; Shih, S. R.; Ho, M. C.; Tai, H. C.; Chang, K. J.; Chen, A.; Chen, C. N., Multi-Reader Multi-Case Study for Performance Evaluation of High-Risk Thyroid Ultrasound with Computer-Aided Detection. *Cancers (Basel)* **2020,** *12* (2), 1-13.

253. Sultan, L. R.; Xiong, H.; Zafar, H. M.; Schultz, S. M.; Langer, J. E.; Sehgal, C. M., Vascularity Assessment of Thyroid Nodules by Quantitative Color Doppler Ultrasound. *Ultrasound in Medicine & Biology* **2015,** *41* (5), 1287-1293.

254. Wu, M.-H.; Chen, C.-N.; Chen, K.-Y.; Ho, M.-C.; Tai, H.-C.; Chung, Y.-C.; Lo, C.-P.; Chen, A.; Chang, K.-J.; M.-H, W.; C.-N, C.; K.-Y, C.; M.-C, H.; H.-C, T.; Y.-C, C.; C.-P, L.; A, C.; K.-J, C., Quantitative Analysis of Dynamic Power Doppler Sonograms for Patients with Thyroid Nodules. *Ultrasound in Medicine & Biology* **2013,** *39* (9), 1543-1551.

255. Lyshchik, A.; Moses, R.; Barnes, S. L.; Higashi, T.; Asato, R.; Miga, M. I.; Gore, J. C.; Fleischer, A. C., Quantitative Analysis of Tumor Vascularity in Benign and Malignant Solid Thyroid Nodules. *Journal of Ultrasound in Medicine* **2007,** *26* (6), 837-846.

256. Gharib, H.; Papini, E.; Garber, J. R.; null, n., American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici

Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules–2016 update. *Endocr Pract* **2016,** *22*, 622.

257. Shin, J. H.; Baek, J. H.; Chung, J.; null, n., Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. *Korean J Radiol* **2016,** *17*, 370.

258. Choi, Y. J.; Baek, J. H.; Baek, S. H., Web-based malignancy risk estimation for thyroid nodules using ultrasonography characteristics: development and validation of a predictive model. *Thyroid* **2015,** *25*, 1306.

259. Ha, S. M.; Ahn, H. S.; Baek, J. H.; Ahn, H. Y.; Chung, Y. J.; Cho, B. Y.; Park, S. B., Validation of Three Scoring Risk-Stratification Models for Thyroid Nodules. *Thyroid* **2017,** *27* (12), 1550-1557.

260. Sollini, M.; Cozzi, L.; Chiti, A.; Kirienko, M., Texture analysis and machine learning to characterize suspected thyroid nodules and differentiated thyroid cancer: Where do we stand? *European journal of radiology* **2018,** *99*, 1-8.

261. Yu, Q.; Jiang, T.; Zhou, A.; Zhang, L.; Zhang, C.; Xu, P., Computer-aided diagnosis of malignant or benign thyroid nodes based on ultrasound images. *Eur Arch Otorhinolaryngol* **2017,** *274* (7), 2891-2897.

262. Lu, Y.; Shi, X. Q.; Zhao, X.; Song, D.; Li, J., Value of Computer Software for Assisting Sonographers in the Diagnosis of Thyroid Imaging Reporting and Data System Grade 3 and 4 Thyroid Space-Occupying Lesions. *Journal of Ultrasound in Medicine* **2019,** *38* (12), 3291-3300.

263. Chambara, N.; Liu, S. Y. W.; Lo, X.; Ying, M., Diagnostic performance evaluation of different TI-RADS using ultrasound computer-aided diagnosis of thyroid nodules: An experience with adjusted settings. *PLoS One* **2021,** *16* (1), e0245617.

264. Chambara, N.; Liu, S. Y. W.; Lo, X.; Ying, M., Comparative Analysis of Computer-Aided Diagnosis and Computer-Assisted Subjective Assessment in Thyroid Ultrasound. *Life* **2021,** *11* (11), 1148.

265. Pedraza, L.; Vargas, C.; Narváez, F.; Durán, O.; Muñoz, E.; Romero, E., An open access thyroid ultrasound-image Database. In *10th International Symposium on Medical Information Processing and Analysis*, 2015; Vol. 9287.

266. AmCAD BioMed Corporation, AmCAD-UT Detection User's Guide AmCAD BioMed Corporation

Taipei, Taiwan, 2015; Vol. Version 2.2, pp 53-61.

267. Seo, H.; Na, D. G.; Kim, J.-H. H.; Kim, K. W.; Yoon, J. W., Ultrasound-Based Risk Stratification for Malignancy in Thyroid Nodules: A Four-Tier Categorization System. *European Radiology* **2015,** *25* (7), 2153-2162.

268. Moon, H. J.; Kwak, J. Y.; Kim, E. K.; Kim, M. J., A taller-than-wide shape in thyroid nodules in transverse and longitudinal ultrasonographic planes and the prediction of malignancy. *Thyroid* **2011,** *21* (11), 1249-53.

269. Grani, G.; Lamartina, L.; Ramundo, V.; Falcone, R.; Lomonaco, C.; Ciotti, L.; Barone, M.; Maranghi, M.; Cantisani, V.; Filetti, S.; Durante, C., Taller-Than-Wide Shape: A New Definition Improves the Specificity of TIRADS Systems. *Eur Thyroid J* **2020,** *9* (2), 85-91.

270. Saito, T.; Rehmsmeier, M., The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE* **2015,** *10* (3), e0118432.

271. Habibzadeh, F.; Habibzadeh, P.; Yadollahie, M., On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochem Med (Zagreb)* **2016,** *26* (3), 297-307.

272. Unal, I., Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Computational and Mathematical Methods in Medicine* **2017,** *2017*, 3762651.

273. Hinkle, D. E.; Wiersma, W.; Jurs, S. G., Correlation: A measure of relationship. *Applied statistics for the behavioral sciences, 5th ed. Boston: Houghton Mifflin* **2003**, 95-120.

274. Landis, J. R.; Koch, G. G., The measurement of observer agreement for categorical data. *Biometrics* **1977,** *33* (1), 159-74.

275. Shen, Y.; Liu, M.; He, J.; Wu, S.; Chen, M.; Wan, Y.; Gao, L.; Cai, X.; Ding, J.; Fu, X., Comparison of Different Risk-Stratification Systems for the Diagnosis of Benign and Malignant Thyroid Nodules. *Frontiers in Oncology* **2019,** *9* (May), 1-8.

276. Zhang, W. B.; Xu, H. X.; Zhang, Y. F.; Guo, L. H.; Xu, S. H.; Zhao, C. K.; Liu, B. J., Comparisons of ACR TI-RADS, ATA guidelines, Kwak TI-RADS, and KTA/KSThR guidelines in malignancy risk stratification of thyroid nodules. *Clin Hemorheol Microcirc* **2020,** *75* (2), 219-232.

277. Liang, X.-W.; Cai, Y.-Y.; Yu, J.-S.; Liao, J.-Y.; Chen, Z.-Y., Update on thyroid ultrasound: a narrative review from diagnostic criteria to artificial intelligence techniques. *Chinese medical journal* **2019,** *132* (16), 1974-1982.

278. Yoon, S. J.; Na, D. G.; Gwon, H. Y.; Paik, W.; Kim, W. J.; Song, J. S.; Shim, M. S., Similarities and Differences Between Thyroid Imaging Reporting and Data Systems. *American Journal of Roentgenology* **2019,** *213* (2), W76-W84.

279. 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., Thyroid Imaging Reporting and Data System for US Features of Nodules: A Step in Establishing Better Stratification of Cancer Risk. *Radiology* **2011,** *260* (3), 892-899.

280. Yoon, J. H.; Lee, H. S.; Kim, E.-K.; Moon, H. J.; Kwak, J. Y., Malignancy Risk Stratification of Thyroid Nodules: Comparison between the Thyroid Imaging Reporting and

Data System and the 2014 American Thyroid Association Management Guidelines. *Radiology* **2016,** *278* (3), 917-924.

281. Jeong, E. Y.; Kim, H. L.; Ha, E. J.; Park, S. Y.; Cho, Y. J.; Han, M., Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. *European Radiology* **2019,** *29* (4), 1978-1985.

282. Remonti, L. R.; Kramer, C. K.; Leitao, C. B.; Pinto, L. C.; Gross, J. L., Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid* **2015,** *25* (5), 538-50.

283. Ha, S. M.; Chung, Y. J.; Ahn, H. S.; Baek, J. H.; Park, S. B., Echogenic foci in thyroid nodules: diagnostic performance with combination of TIRADS and echogenic foci. *BMC Medical Imaging* **2019,** *19* (1), 28.

284. Thomas, J.; Ledger, G. A.; Mamillapalli, C. K., Use of artificial intelligence and machine learning for estimating malignancy risk of thyroid nodules. *Curr Opin Endocrinol Diabetes Obes* **2020,** *27* (5), 345-350.

285. Park, S. H., Artificial intelligence for ultrasonography: unique opportunities and challenges. *Ultrasonography* **2021,** *40* (1), 3-6.

286. Glas, A. S.; Lijmer, J. G.; Prins, M. H.; Bonsel, G. J.; Bossuyt, P. M., The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* **2003,** *56* (11), 1129-35.

287. Eusebi, P., Diagnostic accuracy measures. *Cerebrovasc Dis* **2013,** *36* (4), 267-72.

288. Russ, G.; Trimboli, P.; Buffet, C., The New Era of TIRADSs to Stratify the Risk of Malignancy of Thyroid Nodules: Strengths, Weaknesses and Pitfalls. *Cancers* **2021,** *13* (17), 4316.

289. Lee, S.; Lee, D. K., What is the proper way to apply the multiple comparison test? *Korean J Anesthesiol* **2018,** *71* (5), 353-360.

290. Słowińska-Klencka, D.; Wysocka-Konieczna, K.; Klencki, M.; Popowicz, B., Diagnostic Value of Six Thyroid Imaging Reporting and Data Systems (TIRADS) in Cytologically Equivocal Thyroid Nodules. *Journal of Clinical Medicine* **2020,** *9* (7), 2281.

291. Słowińska-Klencka, D.; Wysocka-Konieczna, K.; Klencki, M.; Popowicz, B., Usability of EU-TIRADS in the Diagnostics of Hürthle Cell Thyroid Nodules with Equivocal Cytology. *Journal of Clinical Medicine* **2020,** *9* (11), 3410.

292. Wettasinghe, M. C.; Rosairo, S.; Ratnatunga, N.; Wickramasinghe, N. D., Diagnostic accuracy of ultrasound characteristics in the identification of malignant thyroid nodules. *BMC Res Notes* **2019,** *12* (1), 193.

293. Deng, S.; Jiang, Q.; Zhu, Y.; Zhang, Y., An analysis of the clinical value of high-frequency color Doppler ultrasound in the differential diagnosis of benign and malignant thyroid nodules. *Int J Clin Exp Med* **2018,** *11* (3), 2331-2336.

294. Kalantari, S., The Diagnostic Value of Color Doppler Ultrasonography in Predicting Thyroid Nodules Malignancy. *The International Tinnitus Journal* **2018,** *22* (1), 35-39.

295. Chen, L.; Zhan, J.; Diao, X. H.; Liu, Y. C.; Shi, Y. X.; Chen, Y.; Zhan, W. W., Additional Value of Superb Microvascular Imaging for Thyroid Nodule Classification with the Thyroid Imaging Reporting and Data System. *Ultrasound in Medicine and Biology* **2019,** *45* (8), 2040-2048.

296. Ahn, H. S.; Lee, J. B.; Seo, M.; Park, S. H.; Choi, B. I., Distinguishing benign from malignant thyroid nodules using thyroid ultrasonography: utility of adding superb microvascular imaging and elastography. *Radiologia Medica* **2018,** *123* (4), 260-270.

297. Yang, G. C. H.; Fried, K. O., Most Thyroid Cancers Detected by Sonography Lack Intranodular Vascularity on Color Doppler Imaging: Review of the Literature and Sonographic-Pathologic Correlations for 698 Thyroid Neoplasms. *Journal of Ultrasound in Medicine* **2017,** *36* (1), 89-94.

298. Fukunari, N.; Nagahama, M.; Sugino, K.; Mimura, T.; Ito, K.; Ito, K., Clinical Evaluation of Color Doppler Imaging for the Differential Diagnosis of Thyroid Follicular Lesions. *World Journal of Surgery* **2004,** *28* (12), 1261-1265.

299. Fresilli, D.; David, E.; Pacini, P.; Del Gaudio, G.; Dolcetti, V.; Lucarelli, G. T.; Di Leo, N.; Bellini, M. I.; D'Andrea, V.; Sorrenti, S.; Mascagni, D.; Biffoni, M.; Durante, C.; Grani, G.; De Vincentis, G.; Cantisani, V., Thyroid Nodule Characterization: How to Assess the Malignancy Risk. Update of the Literature. *Diagnostics (Basel)* **2021,** *11* (8), 1374.

300. Nattabi, H. A.; Sharif, N. M.; Yahya, N.; Ahmad, R.; Mohamad, M.; Zaki, F. M.; Yusoff, A. N., Is Diagnostic Performance of Quantitative 2D-Shear Wave Elastography Optimal for Clinical Classification of Benign and Malignant Thyroid Nodules?. A Systematic Review and Meta-analysis. *Academic Radiology* **2017,** *29* (3), S114-S121.

301. Swan, K. Z.; Nielsen, V. E.; Bonnema, S. J., Evaluation of thyroid nodules by shear wave elastography: a review of current knowledge. *Journal of Endocrinological Investigation* **2021,** *44* (10), 2043-2056.

302. Park, A. Y.; Son, E. J.; Han, K.; Youk, J. H.; Kim, J. A.; Park, C. S., Shear wave elastography of thyroid nodules for the prediction of malignancy in a large scale study. *Eur J Radiol* **2015,** *84* (3), 407-412.

303. Han, R.-J.; Du, J.; Li, F.-H.; Zong, H.-R.; Wang, J.-D.; Shen, Y.-L.; Zhou, Q.-Y., Comparisons and Combined Application of Two-Dimensional and Three-Dimensional Real-time Shear Wave Elastography in Diagnosis of Thyroid Nodules. *J Cancer* **2019,** *10* (9), 1975-1984.

304. Cantisani, V.; David, E.; Grazhdani, H.; Rubini, A.; Radzina, M.; Dietrich, C. F.; Durante, C.; Lamartina, L.; Grani, G.; Valeria, A.; Bosco, D.; Di Gioia, C.; Frattaroli, F. M.; D'Andrea, V.; De Vito, C.; Fresilli, D.; D'Ambrosio, F.; Giacomelli, L.; Catalano, C., Prospective Evaluation of Semiquantitative Strain Ratio and Quantitative 2D Ultrasound

ShearWave Elastography (SWE) in Association with TIRADS Classification for Thyroid Nodule Characterization. *Ultraschall in Der Medizin* **2019,** *40* (4), 495-503.

305. Dobruch-Sobczak, K.; Zalewska, E. B.; Gumińska, A.; Słapa, R. Z.; Mlosek, K.; Wareluk, P.; Jakubowski, W.; Dedecjus, M., Diagnostic Performance of Shear Wave Elastography Parameters Alone and in Combination with Conventional B-Mode Ultrasound Parameters for the Characterization of Thyroid Nodules: A Prospective, Dual-Center Study. *Ultrasound in Medicine & Biology* **2016,** *42* (12), 2803-2811.

306. Gangadhar, K.; Hippe, D. S.; Thiel, J.; Dighe, M., Impact of Image Orientation on Measurements of Thyroid Nodule Stiffness Using Shear Wave Elastography. *Journal of Ultrasound in Medicine* **2016,** *35* (8), 1661-1667.

307. Dighe, M.; Hippe, D. S.; Thiel, J., Artifacts in Shear Wave Elastography Images of Thyroid Nodules. *Ultrasound in medicine & biology* **2018,** *44* (6), 1170-1176.

308. Swan, K. Z.; Bonnema, S. J.; Jespersen, M. L.; Nielsen, V. E., Reappraisal of shear wave elastography as a diagnostic tool for identifying thyroid carcinoma. *Endocrine connections* **2019,** *8* (8), 1195-1205.

309. Russ, G.; Royer, B.; Bigorgne, C.; Rouxel, A.; Bienvenu-Perrard, M.; Leenhardt, L., Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. *Eur J Endocrinol* **2013,** *168* (5), 649-55.

310. Wang, F.; Chang, C.; Gao, Y.; Chen, Y. L.; Chen, M.; Feng, L. Q., Does shear wave elastography provide additional value in the evaluation of thyroid nodules that are suspicious for malignancy? *Journal of Ultrasound in Medicine* **2016,** *35* (11), 2397-2404.

311. Liu, Z.; Jing, H.; Han, X.; Shao, H.; Sun, Y.-X.; Wang, Q.-C.; Cheng, W., Shear wave elastography combined with the thyroid imaging reporting and data system for malignancy risk stratification in thyroid nodules. *Oncotarget* **2017,** *8* (26), 43406-43416.

312. Yeon, E. K.; Sohn, Y. M.; Seo, M.; Kim, E. J.; Eun, Y. G.; Park, W. S.; Yun, S. J., Diagnostic Performance of a Combination of Shear Wave Elastography and B-Mode Ultrasonography in Differentiating Benign From Malignant Thyroid Nodules. *Clin Exp Otorhinolaryngol* **2020,** *13* (2), 186-193.

313. Kim, H.; Kim, J. A.; Son, E. J.; Youk, J. H., Quantitative assessment of shear-wave ultrasound elastography in thyroid nodules: Diagnostic performance for predicting malignancy. *European Radiology* **2013,** *23* (9), 2532-2537.

314. Yoo, M. H.; Kim, H. J.; Choi, I. H.; Park, S.; Kim, S. J.; Park, H. K.; Byun, D. W.; Suh, K., Shear wave elasticity by tracing total nodule showed high reproducibility and concordance with fibrosis in thyroid cancer. *BMC Cancer* **2020,** *20* (1), 118.

315. Bhatia, K. S. S.; Tong, C. S. L.; Cho, C. C. M.; Yuen, E. H. Y.; Lee, Y. Y. P.; Ahuja, A. T., Shear wave elastography of thyroid nodules in routine clinical practice: Preliminary observations and utility for detecting malignancy. *European Radiology* **2012,** *22* (11), 2397-2406.

316. Wang, F.; Chang, C.; Chen, M.; Gao, Y.; Chen, Y.-L. L.; Zhou, S.-C. C.; Li, J.-W. W.; Zhi, W.-X. X., Does lesion size affect the value of shear wave elastography for differentiating between benign and malignant thyroid nodules? *Journal of Ultrasound in Medicine* **2018,** *37* (3), 601-609.

317. He, Y. P.; Xu, H. X.; Wang, D.; Li, X. L.; Ren, W. W.; Zhao, C. K.; Bo, X. W.; Liu, B. J.; Yue, W. W., First experience of comparisons between two different shear wave speed imaging systems in differentiating malignant from benign thyroid nodules. *Clin Hemorheol Microcirc* **2017,** *65* (4), 349-361.

318. Yang, Q.; Zhou, W. H.; Li, J. Y.; Wu, G. J.; Ding, F.; Tian, X. S., Comparative Analysis of Diagnostic Value for Shear Wave Elastography and Real-Time Elastographic Imaging for Thyroid Nodules. *J. Med. Imaging Health Inform.* **2019,** *9* (2), 334-338.

319. Veyrieres, J. B.; Albarel, F.; Lombard, J. V.; Berbis, J.; Sebag, F.; Oliver, C.; Petit, P., A threshold value in Shear Wave elastography to rule out malignant thyroid nodules: A reality? *European Journal of Radiology* **2012,** *81* (12), 3965-3972.

320. Liu, B.; Liang, J.; Zheng, Y.; Xie, X.; Huang, G.; Zhou, L.; Wang, W.; Lu, M., Two-dimensional shear wave elastography as promising diagnostic tool for predicting malignant thyroid nodules: a prospective single-centre experience. *Eur Radiol* **2015,** *25* (3), 624-34.

321. Tan, S.; Sun, P.-F.; Xue, H.; Fu, S.; Zhang, Z.-P.; Mei, F.; Miao, L.-Y.; Wang, X.-H., Evaluation of thyroid micro-carcinoma using shear wave elastography: Initial experience with qualitative and quantitative analysis. *European Journal of Radiology* **2021,** *137*, 109571.

322. Fukuhara, T.; Matsuda, E.; Endo, Y.; Takenobu, M.; Izawa, S.; Fujiwara, K.; Kitano, H., Correlation between quantitative shear wave elastography and pathologic structures of thyroid lesions. *Ultrasound in medicine & biology* **2015,** *41* (9), 2326-2332.

323. Szczepanek-Parulska, E.; Wolin´ski, K.; Wolin´ski, W.; Stangierski, A.; Gurgul, E.; Ruchała, M.; Woliński, K.; Stangierski, A.; Gurgul, E.; Ruchała, M., Biochemical and ultrasonographic parameters influencing thyroid nodules elasticity. *Endocrine* **2014,** *47*, 519-527.

324. Li, H.; Kang, C.; Xue, J.; Jing, L.; Miao, J., Influence of lesion size on shear wave elastography in the diagnosis of benign and malignant thyroid nodules. *Scientific Reports* **2021,** *11* (1).

325. Chanda, A.; Callaway, C., Tissue Anisotropy Modeling Using Soft Composite Materials. *Applied Bionics and Biomechanics* **2018,** *2018*, 4838157.

326. Zhao, C.-K.; Chen, S.-G.; Alizad, A.; He, Y.-P.; Wang, Q.; Wang, D.; Yue, W.-W.; Zhang, K.; Qu, S.; Wei, Q.; Xu, H.-X., Three-Dimensional Shear Wave Elastography for

Differentiating Benign From Malignant Thyroid Nodules. *Journal of Ultrasound in Medicine* **2018,** *37* (7), 1777-1788.

327. Fukuhara, T.; Matsuda, E.; Endo, Y.; Donishi, R.; Izawa, S.; Fujiwara, K.; Kitano, H.; Takeuchi, H., Impact of Fibrotic Tissue on Shear Wave Velocity in Thyroid: An Ex Vivo Study with Fresh Thyroid Specimens. *BioMed research international* **2015,** *2015*, 569367-569367.

328. Yi, L.; Qiong, W.; Yan, W.; Youben, F.; Bing, H., Correlation between Ultrasound Elastography and Histologic Characteristics of Papillary Thyroid Carcinoma. *Sci Rep* **2017,** *7*, 45042.

329. Haugen, B. R., 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: What is new and what has changed? *Cancer* **2017,** *123* (3), 372-381.

330. Azizi, G.; Keller, J. M.; Mayo, M. L.; Piper, K.; Puett, D.; Earp, K. M.; Malchoff, C. D., Shear wave elastography and Afirma™ gene expression classifier in thyroid nodules with indeterminate cytology: a comparison study. *Endocrine* **2018,** *59* (3), 573-584.

331. Marturano, I.; Russo, M.; Malandrino, P.; Buscema, M.; La Rosa, G. L.; Spadaro, A.; Manzella, L.; Sciacca, L.; L'Abbate, L.; Rizzo, L., Combined use of sonographic and elastosonographic parameters can improve the diagnostic accuracy in thyroid nodules at risk of malignancy at cytological examination. *Minerva Endocrinol* **2020,** *45* (1), 3-11.

332. Celletti, I.; Fresilli, D.; De Vito, C.; Bononi, M.; Cardaccio, S.; Cozzolino, A.; Durante, C.; Grani, G.; Grimaldi, G.; Isidori, A. M.; Catalano, C.; Cantisani, V., TIRADS, SRE and SWE in INDETERMINATE thyroid nodule characterization: Which has better diagnostic performance? *La radiologia medica* **2021,** *126* (9), 1189-1200.

333. Zhang, W.-B.; Li, J.-J.; Chen, X.-Y.; He, B.-L.; Shen, R.-H.; Liu, H.; Chen, J.; He, X.-F., SWE combined with ACR TI-RADS categories for malignancy risk stratification of

thyroid nodules with indeterminate FNA cytology. *Clinical Hemorheology and Microcirculation* **2020,** *76*, 381-390.

334. Liu, J.; Chen, Y.; Zheng, Y.; Wen, D.; Wang, Y.; Xue, G., The role of SWE and ATA (2015) guidelines combined mode in differentiation malignant from benign of Bethesda III thyroid nodules. *Lin chuang er bi yan hou tou jing wai ke za zhi= Journal of clinical otorhinolaryngology, head, and neck surgery* **2018,** *32* (18), 1400-1405.

Appendix 1

Proportion of agreement (%) between raters based on TIRADS cut-off points for malignancy-risk stratification

Raters	Nodules	TIRADS			
		AACE (%)	ATA(%)	EU (%)	KSThR(%)
R_1 vs CAD	M	84.9	89.7	77.4	87.1
	B	63.9	87.7	71.0	73.0
	ALL	73.7	88.5	73.5	78.4
R_2 vs CAD	M	79.2	84.6	79.0	82.3
	B	50.8	73.7	63.0	61.0
	ALL	64.0	78.1	69.1	69.1
R_1 vs R_2	M	86.8	89.7	79.0	85.4
	B	80.3	87.7	80.0	72
	ALL	83.3	88.5	79.6	77.2

M = malignant, B = benign

Appendix 2

Prevalence-based diagnostic performance outcomes with sensitivity and specificity for common specified nodules across different TIRADS

RATER BY TIRADS	TOTAL NODULES	SEN % (CI)	SPE (%) (CI)	PPV (CI)	NPV (CI)	DA (CI)
EU-CAD	162	79.0 (66.8;88.3)	55.0 (44.7;65.0)	52.1 (41.6;62.5)	80.9 (69.5;89.4)	64.2 (56.3;71.6)
EU-R_1	162	85.5 (74.2;93.1)	62.0 (51.8;71.5)	58.2 (47.4;68.5)	87.3 (77.3;94.1)	71.0 (63.3;77.8)
EU-R_2	162	71.0 (58.1;81.8)	64.0 (53.8;73.4)	55.0 (43.5;66.2)	78.1 (67.6;86.4)	66.7 (58.8;73.9)
KSThR- CAD	162	83.9 (72.3;92.0)	46.0 (36.0;56.3)	49.1 (39.2;59.0)	82.1 (69.6;91.1)	60.5 (52.5;68.1)
KSThR-R_1	162	90.3 (80.1;96.4)	51.0 (40.8;61.4)	53.3 (43.3;63.1)	89.5 (78.5;96.0)	66.1 (58.2;73.3)
KSThR-R_2	162	75.8 (63.3;85.8)	61.0 (50.7;70.6)	54.7 (43.6;65.4)	80.3 (69.6;88.5)	66.7 (58.8;73.9)
AACE-CAD	114	92.5 (81.8;97.9)	26.2 (15.8;39.1)	52.1 (41.6;62.5)	80.0 (56.3;94.3)	57.0 (47.4;62.5)
AACE-R1	114	88.7 (77.0;95.7)	54.1 (40.9;66.9)	62.7 (50.7;73.6)	84.6 (69.5;94.1)	70.2 (60.9;78.4)
AACE-R_2	114	79.3 (65.9;89.2)	62.3 (49.0;74.4)	64.6 (51.8;76.1)	77.6 (63.4;88.3)	70.2 (60.9;78.4)
ATA-CAD	96	79.5 (63.5;90.7)	66.7 (52.9;78.6)	62.0 (47.2;75.4)	82.6 (68.6;92.2)	71.9 (61.8;80.6)
ATA-R_1	96	79.5 (63.5;90.7)	70.2 (56.6;81.6)	64.6 (49.5;77.8)	83.3 (69.8;92.5)	74.0 (64.0;82.4)
ATA-R_2	96	74.4 (57.9;87.0)	68.4 (54.8;80.1)	61.7 (46.4;75.5)	79.6 (65.7;89.8)	70.8 (60.7;79.7)

NPV = negative predictive value, PPV = positive predictive value, DA = diagnostic accuracy

Appendix 3

p- values for SEN, SPEC and AUROC comparisons between subjective raters and CAD per TIRADS

TIRADS	Diagnostic Performance Measure	Raters		
		R_1 vs CAD	R_2 vs CAD	R_1 vs R_2
EU	SEN	0.82	0.51	0.041
	SPEC	0.27	0.19	0.82
	PPV	0.40	0.70	0.67
	NPV	0.30	0.67	0.13
	DA	0.54	0.066	0.16
	AUROC	0.064	0.57	0.11
KSThR	SEN	0.57	0.31	0.010
	SPEC	0.21	0.020	0.37
	PPV	0.54	0.44	0.86
	NPV	0.26	0.79	0.15
	DA	0.74	0.70	0.23
	AUROC	0.021	0.13	0.41
AACE	SEN	0.50	0.017	0.089
	SPEC	0.001	<0.001	0.81
	PPV	0.17	0.12	0.81
	NPV	0.65	0.83	0.41
	DA	0.001	<0.001	0.30
	AUROC	0.001	<0.001	0.20
ATA	SEN	0.99	0.63	0.63
	SPEC	0.73	0.99	0.99
	PPV	0.79	0.98	0.77
	NPV	0.93	0.71	0.64
	DA	0.78	0.42	0.76
	AUROC	0.82	0.95	0.72