Aus der Klinik für Radiologie und Nuklearmedizin
der Medizinischen Fakultät
der Otto-von-Guericke-Universität Magdeburg
INKA- Application Driven Research

A Systematic Approach to the Detection, Quantification and Classification of Thyroid Nodules in Ultrasound Images using Image Computing, Machine and Deep Learning for Reduced Subjectivity and Inter- and Intraobserver Variability

D i s s e r t a t i o n

zur Erlangung des Doktorgrades
Dr. rer. medic.
(doctor rerum medicarum)

an der Medizinischen Fakultät
der Otto-von-Guericke-Universität Magdeburg

vorgelegt von: Elmer Jeto Gomes Ataide
aus Portugal
Magdeburg 2023

Supervised by:

Prof. Dr. rer. medic. Michael Friebe

Prof. Dr. med. Michael Kreissl

Dr. Alfredo Illanes

Dr. med. Simone Schenke

"Anyone can learn at any time. Be it from an adult or a child. As long as you keep an open mind."

- M.A

Bibliographische Beschreibung:

Gomes Ataide, Elmer Jeto:

A Systematic Approach to the Detection, Quantification and Classification of Thyroid Nodules in Ultrasound Images using Image Computing, Machine and Deep Learning for Reduced Subjectivity and Inter- and Intraobserver Variability - 2023. - 121 Bl(pages)., 9 Abb., 2 Tab., 1 Anl.

Abstract

Ultrasound imaging is used as a first and most frequent mode for assessing thyroid nodules. Nodule features like irregular shape, microcalcifications, and taller-than-wide morphology raise suspicion for malignancy. However, classification of thyroid nodules using ultrasound images depends heavily on the interpreting physician's experience and skill, leading to inherent issues with subjectivity and both interobserver and intraobserver variability. This thesis presents a systematic approach for the detection, region estimation and classification of thyroid nodules using ultrasound images aimed at reducing overall subjectivity and inter- and intraobserver variability. This is achieved through the use of texture analysis, feature extraction, machine learning and deep learning using ultrasound images with thyroid nodules. The study begins by differentiating textures of the thyroid gland from surrounding organs in US images using autoregressive features and machine learning. This is followed by the comparison of performances from four different deep learning algorithms for the detection and localization of thyroid nodules. Subsequently, we estimate and quantify the solid and cystic regions within thyroid nodules using textural analysis and machine learning. Lastly, extraction of geometric and morphological features, helps classify thyroid nodules using machine learning techniques that consider the visual characteristics analyzed by physicians according to TIRADS. Thus, providing them with quantifiable evidence that supports the classification process. Future research will focus on refining

and validating these models, and determine how best to incorporate them into existing clinical workflows.

Keywords

Ultrasound Imaging, Thyroid gland, Thyroid Nodules, Data Processing, Image computing, Texture Analysis, Feature Extraction, Machine Learning, Deep Learning, Segmentation, Classification.

Table of Contents

Number	Title	Page Number
	List of Figures and Tables	
	List of Abbreviations	
	Preamble, List of Publications	
1	Introduction	1
1.1	The Thyroid Gland	1
1.2	Diseases Affecting the Thyroid Gland	2
1.3	Thyroid Nodules	3
1.3.1	Types of Thyroid Nodules	4
1.4	Diagnosis of Thyroid Nodules using Medical Imaging	6
1.5	Diagnosis of Thyroid Nodules using Ultrasound Imaging	8
1.5.1	Thyroid Imaging Reporting and Data System (TIRADS)	9
1.6	Technical State-of-the-art	10
1.6.1	Texture Analysis	10
1.6.2	Machine Learning	12
1.6.3	Deep Learning	14
1.6.4	Shape-based Features	16
1.6.5	Application of Discussed State of the Art Technologies in Thyroid Ultrasound Images	17
1.7	Motivation and Contribution	22
1.7.1	Motivation	22

1.7.2	Contribution	23
2	Hypotheses	27
2.1	Methodology	28
2.1.1	Detection of the Thyroid Gland using Features Extracted from US Images and Classification using Machine Learning for the Localization of the Thyroid Gland (H1-Publication 1)	28
2.1.2	A Comparison of Deep Learning Architectures for the Detection of Thyroid Nodules in US Images (H2 - Publication 2)	30
2.1.3	Solid and Cystic Region Estimation in Thyroid Nodules using US Images to Facilitate Objective Decision Making Among Physicians through Quantifiable Evidence (H3 - Publication 3)	33
2.1.4	Classification of Thyroid Nodules in US Images using Shape based Features and Machine Learning to Support Clinical Decisions with Currently Considered Visual Characteristics (H4 - Publication 4)	34
3	Results	36
3.1	Results in relation to Publication 1	36
3.2	Results in relation to Publication 2	36
3.3	Results in relation to Publication 3	37
3.4	Results in relation to Publication 4	39
4	Discussion, Conclusion and Future Scope	41
4.1	Limitations and Future Scope	45
	Summary	47
	Zusammenfassung	49
	Bibliography	

	Danksagung	
	Eidesstattiche Erklärung	
	Description of educational background	
	List of All Publications	
	Publications included in the Dissertation	

List of Figures and Tables

Number	Title	Page Number
Figure 1	Schematic diagram of the anatomy of the thyroid gland including the location and parts	1
Figure 2	Examples of thyroid nodules as seen in ultrasound images discovered during routine medical examinations	6
Figure 3	Flow diagram of proposed approach. The boxes in the middle (in blue) highlight the general path followed for the work according to the publications. The boxes (in green) at the top represent the specific feature extraction, machine and deep learning techniques used in each publication while those at the bottom provide a reasoning for the work in contribution to the systematic approach	26

Figure 4	Extraction of 20x20 texture patches from the thyroid gland and surrounding structures collected for the feature extraction	29
Figure 5	Conversion of the extracted texture patches using four different transformations of matrix to a signal. (i) ZigZag, (ii)ZigZag rotated by 90°, (iii) Spiral and (iv) Spiral rotated by 90°	29
Figure 6	U-Net architecture as proposed by Ronnenberger et al. [26] that is used for the segmentation of structures in biomedical images	31
Figure 7	SUMNet architecture as proposed by Nandamuri et al. [67] developed specifically for the segmentation of structures in ultrasound images	31
Figure 8	ResUnet architecture as proposed by Diakogiannis et al. [66] used for the segmentation of images and uses a residual layer together with the architecture of the U-Net	32
Figure 9	Attention U-Net architecture as proposed by Oktay et al. [67] that helps in segmentation tasks by combining a U-Net architecture with attention gates	33

Table 1	Comparison of performance metrics obtained using four different deep learning algorithms for the segmentation of thyroid nodules in US images	37
Table 2	Comparison of performance metrics obtained using three different machine learning algorithms together with two types of extracted features from the same texture database for the estimation of solid and cystic textures within thyroid nodules from US images	39

List of Abbreviations

The abbreviations presented here are in the order of appearance in the dissertation

Number	Full Form	Abbreviation
1	Thyroxine	T4
2	Triiodothyronine	T3
3	Ultrasound	US
4	Fine Needle Aspiration/ Fine Needle Aspiration Biopsy	FNA/FNAB
5	Computed Tomography	CT
6	Magnetic Resonance Imaging	MRI
7	Positron Emission Tomography	PET
8	Thyroid Imaging Reporting and Data System	TIRADS
9	American College of Radiology	ACR
10	American Thyroid Association	ATA
11	State-of-the-Art	SOTA
12	Machine Learning	ML
13	Deep Learning	DL
14	Grey Level Co-Occurrence Matrix	GLCM
15	Grey Level Run Length Matrix	GLRLM
16	Local Binary Patterns	LBP

17	Support Vector Machines	SVM
18	Random Forest Classifiers	RFC
19	k-Nearest Neighbors	k-NN
20	Artificial Neural Networks	ANN
21	Area Under the Receiver Operating Characteristic Curve	AUC-ROC
22	Convolutional Neural Networks	CNNs
23	Geometric and Morphological Features	G-M
24	Decision Trees	DTs
25	2-Dimensional	2D
26	3-Dimensional	3D
27	Radial Basis Function	RBF
28	Active Contours Without Edges	ACWE
29	Autoregressive	AR
30	Cascaded Convolutional Neural Network	CCNN
31	Multi-Layer Perceptron	MLP
32	Continuous Wavelet Transform	CWT
33	Dice Coefficient/ Dice Scores	DC/ DSC
34	Intersection over Union	IoU

35	Sensitivity	SE
36	Specificity	SP
37	Confusion Matrix	CM
38	Echogenicity-based Quantization	EBQ
39	Joint Classification Regression	JCR
40	Feedforward Neural Network	FNN
41	Graph Cut	GC
42	Pixel-based Classifier	PBC

Preamble

The presented dissertation refers to the publications listed below.
It corresponds with the implementation provisions of the doctoral regulations for doctor rerum medicarum (Dr. rer. medic.) of the Medical Faculty at Otto-von-Guericke-University Magdeburg for a cumulative dissertation in the version dated May 14, 2018.

List of Publications included in the Dissertation

Publication 1

Poudel, Prabal, Alfredo Illanes, **Elmer J. G. Ataide**, Nazila Esmaeili, Sathish Balakrishnan, and Michael Friebe. 2019. "Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction With Machine Learning Approaches." *IEEE Access* 7: 79354–65. 10.1109/ACCESS.2019.2923547

Publication 2

Ataide, Elmer Jeto Gomes, Shubham Agrawal, Aishwarya Jauhari, Axel Boese, Alfredol Illanes, Simone Schenke, Michael C. Kreissl, and Michael Friebe. 2021. "Comparison of Deep Learning Algorithms for Semantic Segmentation of Ultrasound Thyroid Nodules." *Current Directions in Biomedical Engineering* 7 (2): 879–82. https://doi.org/10.1515/cdbme-2021-2224

Publication 3

Ataide, Elmer Jeto Gomes, Mathews S. Jabaraj, Alfredo Illanes, Simone Schenke, Axel Boese, Michael C. Kreissl, and Michael Friebe. 2022. "Thyroid Nodule Region Estimation Using Auto-Regressive Modelling and Machine Learning." *Current*

Directions in Biomedical Engineering 8 (2): 588–91. https://doi.org/10.1515/cdbme-2022-1150

Publication 4

Gomes Ataide, Elmer Jeto, Nikhila Ponugoti, Alfredo Illanes, Simone Schenke, Michael Kreissl, and Michael Friebe. 2020. "Thyroid Nodule Classification for Physician Decision Support Using Machine Learning-Evaluated Geometric and Morphological Features." *Sensors* 20 (21). https://doi.org/10.3390/s20216110.

Chapter 1: Introduction

1.1 The Thyroid Gland

The thyroid gland is a small, butterfly-shaped organ located in the front of the neck, below the larynx and wrapped around the trachea (Figure 1). The gland is an essential part of the endocrine system and plays a pivotal role in the metabolism, growth, and development of the human body.

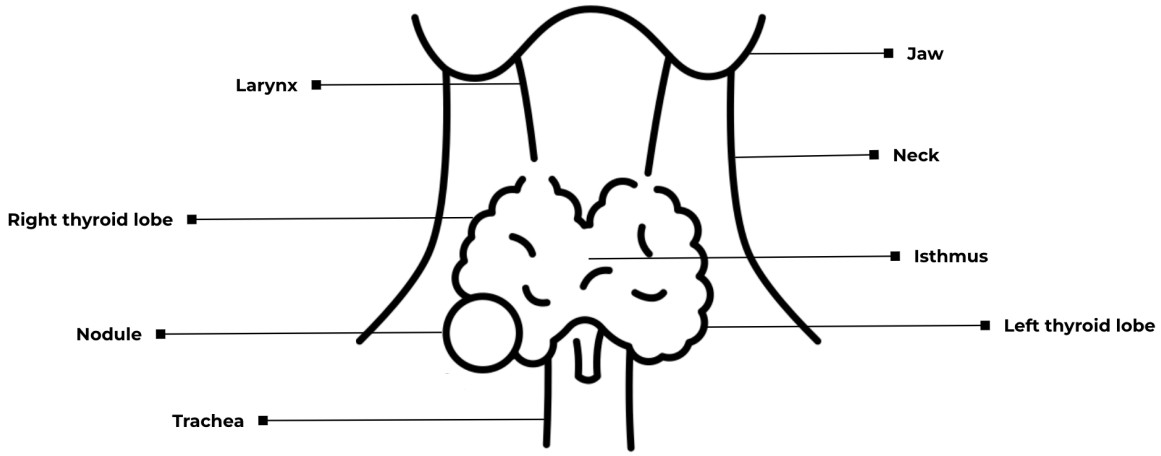


Figure 1: Schematic diagram of the anatomy of the thyroid gland including the location and parts

The primary function of the thyroid gland is to produce, store, and release thyroid hormones into the bloodstream. These hormones include:

1. Thyroxine (T4): The primary hormone produced by the thyroid gland. T4 is an inactive hormone that is converted to its active form, T3, in the liver and other tissues.
2. Triiodothyronine (T3): The active form of thyroid hormone, T3 influences the activity of virtually all the cells and tissues of the body.

3. Calcitonin: Produced by the parafollicular cells (also known as C cells) of the thyroid gland, this hormone helps regulate calcium and phosphate levels in the blood.

The gland is involved in the regulation of metabolism, growth and development, cardiovascular, digestive and reproductive functions. Though small in size, it plays an enormous role in maintaining overall body homeostasis and well-being. Thyroid disorders, therefore, which result in the overproduction (hyperthyroidism) or underproduction (hypothyroidism) of thyroid hormones, can have significant effects on a person's health. [1]

1.2 Diseases Affecting the Thyroid Gland

The thyroid gland can be affected by various conditions that disrupt its normal functioning, potentially leading to a range of symptoms. Some common diseases that affect the thyroid gland include:

1. Hyperthyroidism: This condition occurs when the thyroid gland produces excessive amounts of thyroid hormone, leading to a state of overactivity in the body's metabolism. Common causes include Graves' disease, toxic multinodular goiter, and toxic adenoma. Symptoms can include rapid heart rate, weight loss, anxiety, irritability, tremors, and heat intolerance.

2. Hypothyroidism: Hypothyroidism results from an hypoactive thyroid gland producing insufficient thyroid hormone. Causes include Hashimoto's thyroiditis (an autoimmune disease), surgical removal of the thyroid, and certain medications. Symptoms often develop gradually and can include fatigue, weight gain, depression, constipation, and cold intolerance.

3. Thyroiditis: This is defined as the inflammation of the thyroid, which can cause both hyper- and hypothyroidism. Subtypes include Hashimoto's thyroiditis, subacute (de Quervain's) thyroiditis, and postpartum thyroiditis.

4. Goiter: A goiter is an enlargement of the thyroid gland that can occur for various reasons, such as iodine deficiency, inflammation, or the presence of nodules. Goiters can be associated with hyperthyroidism, hypothyroidism, or normal thyroid function.

5. Thyroid Nodules: Thyroid nodules are lumps within the thyroid gland. They are common, especially in Germany, and can be caused by various conditions, including iodine deficiency, Hashimoto's thyroiditis, and benign or malignant tumors. Most thyroid nodules are asymptomatic and noncancerous, but some can cause symptoms due to their size or the production of excess thyroid hormone. Nodules can be single or multiple, and their detection often leads to further investigations to rule out malignancy, including ultrasound imaging, fine-needle aspiration, or even surgical biopsy. The recent advancements in ultrasound technology and the use of machine learning algorithms have significantly improved the early detection and risk stratification of these nodules, leading to better management.

6. Thyroid Cancer: Although the majority of thyroid nodules are benign, a small proportion can be cancerous. There are several types of thyroid cancer, including papillary, follicular, medullary, and anaplastic thyroid cancer. The prognosis varies widely depending on the type and stage of the cancer at diagnosis. [2]

1.3 Thyroid Nodules

Thyroid nodules are relatively common and can occur due to various reasons. It's recognized that ionizing radiation poses a risk factor for the development of both

benign and malignant thyroid nodules. The individuals exposed to this radiation may experience an annual growth rate of thyroid nodules at about 2% [4]. Furthermore, the occurrence of malignancy in palpable nodules of thyroids previously exposed to radiation has been recorded to be considerably high, ranging between 20% and 50% [3-5]. Additionally, smoking, alcohol, obesity also contribute to the development of thyroid nodules.

Most thyroid nodules are benign (non-cancerous), but a small percentage can be malignant (cancerous). Epidemiological research indicates a prevalence of up to 65% in the adult population [6-7]. Current guidelines suggest that the malignancy rate of these nodules lies between 7 to 15% [8-9].

1.3.1 Types of Thyroid Nodules:

Thyroid nodules can be categorized into two primary types: neoplastic and non-neoplastic. Neoplastic nodules can further be classified as benign or malignant, with benign neoplastic nodules being either functioning or non-functioning. Conversely, non-neoplastic nodules encompass hyperplastic and inflammatory nodules. Thyroid Nodules may be solitary, multiple, cystic, or solid [10]. Despite the fact that over 90% of detected nodules are typically benign and bear no clinical significance [11, 12], it is crucial to note that thyroid nodules are of clinical importance. This is because around 4.0% to 6.5% of these nodules can potentially be indicative of thyroid cancer [13]; however, the malignancy rate in Germany is considerably lower at around 1.1% [11].

1. Colloid Nodules: Also known as adenomatous nodules, these are the most common type and are typically benign. They can occur as a single nodule (solitary thyroid nodule) or as multiple nodules (multinodular goiter).

2. Thyroid Cysts: These nodules are usually filled with fluid or semi-solid material. They may be benign or, in rare cases, contain malignant cells.

3. Inflammatory Nodules: These occur as a result of thyroiditis, or inflammation of the thyroid gland. This can be due to an autoimmune condition like Hashimoto's thyroiditis.

4. Autonomously Functioning Thyroid Nodules: These nodules produce an excess of thyroid hormone on their own, leading to an overactive thyroid, or hyperthyroidism.

5. Thyroid Cancer: Malignant thyroid nodules can be one of several types of thyroid cancer, including papillary thyroid cancer, follicular thyroid cancer, medullary thyroid cancer, and anaplastic thyroid cancer.

It's important to note that the presence of a thyroid nodule does not necessarily mean an individual has thyroid cancer. Many nodules are asymptomatic and are discovered during routine medical examinations, for example, ultrasound imaging. Figure 2 depicts some examples of thyroid nodules where A) is a nodule that is approximately 10% cystic, isoechoic, smooth, wider than tall and no spots, B) and C) are nodules that are approximately 10-50% cystic, isoechoic, smooth, wider than tall and no spots and D) is a nodule that is approximately 10%-20% cystic, isoechoic, smooth, wider than tall and punctated echoic foci. Some may cause symptoms such as a noticeable lump, neck discomfort, voice changes, or symptoms related to thyroid hormone overproduction. However, the early detection and management of these nodules help prevent complications that may arise.

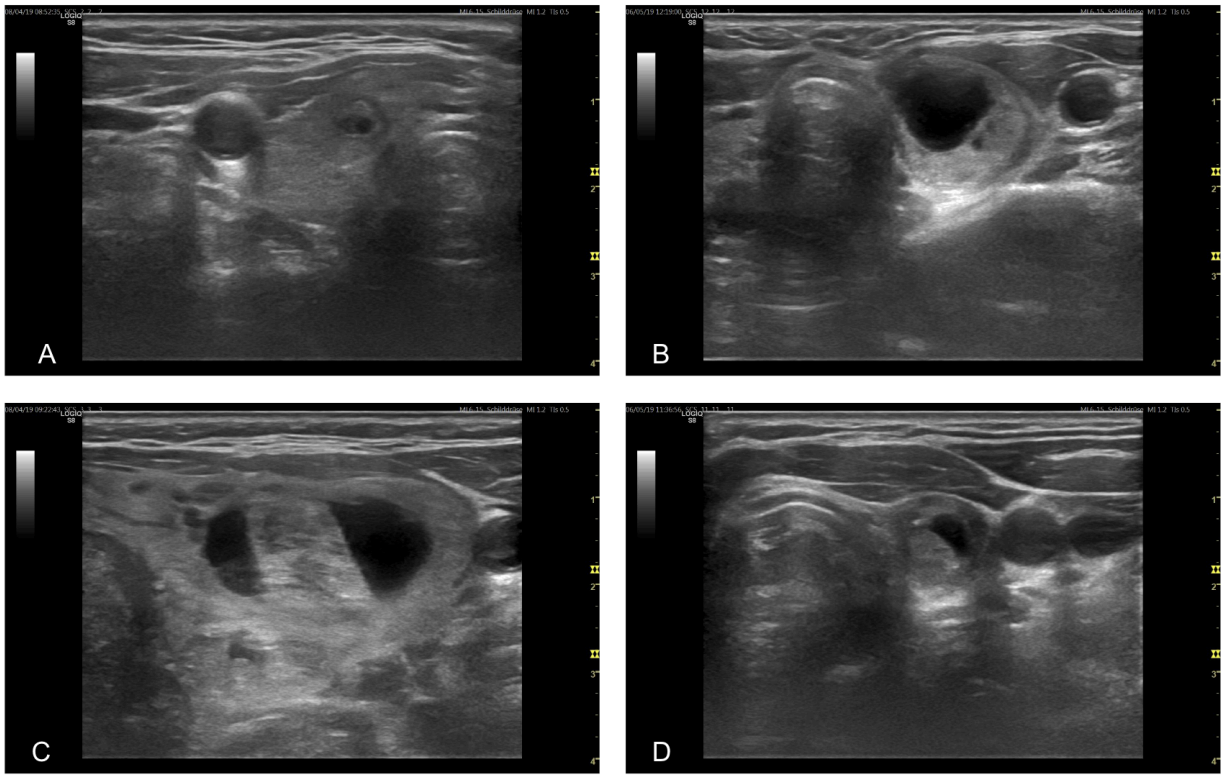


Figure 2: Examples of thyroid nodules as seen in ultrasound images discovered during routine medical examinations.

Diagnostic procedures like ultrasound (US), fine-needle aspiration biopsy, and nuclear medicine imaging can help to determine the type and nature of the nodule, as well as the appropriate treatment strategy, if needed. Treatment can range from watchful waiting in the case of benign, asymptomatic nodules, to surgery, radioactive iodine therapy, or thyroid hormone suppression therapy for more serious conditions [14].

1.4 Diagnosis of Thyroid Nodules using Medical Imaging

The diagnosis of thyroid nodules can be performed using different imaging techniques These are listed below:

1. Ultrasound (US) Imaging: This is the most commonly used imaging modality for the initial evaluation of thyroid nodules. US uses sound waves to create images of the thyroid gland and can provide details about the size, number, and characteristics (such as solid or cystic, smooth or irregular borders) of thyroid nodules. It can also identify if a nodule has characteristics suggestive of malignancy. This imaging modality is non-invasive, does not use radiation, and can also guide a fine-needle aspiration biopsy (FNAB), if needed.

2. Radionuclide Imaging: Also known as thyroid scan, this involves the use of small amounts of radiotracer, typically Tc-99m or I-23 and a special camera to create images of the thyroid gland. It is particularly useful in assessing the functionality of a thyroid nodule. "Hot" nodules, which take up more of the radiotracer material, are usually benign, while "cold" nodules, which take up less radiotracer, can sometimes be malignant.

3. Computed Tomography (CT): While CT is not routinely used to evaluate thyroid nodules due to the radiation exposure, it can be used to assess the extension of large thyroid nodules or goiters into the chest or to evaluate the lymph nodes in the neck for evidence of spread of thyroid cancer.

4. Magnetic Resonance Imaging (MRI): Similar to CT, MRI is not typically used as a first-line for evaluation of thyroid nodules. However, it can provide detailed images of the neck and can be useful in certain cases, particularly for evaluating large goiters or detecting spread of thyroid cancer.

5. Positron Emission Tomography (PET): While not commonly used for initial evaluation of thyroid nodules, PET can be useful in specific situations, such as in the

evaluation of thyroid nodules in patients with a history of malignancy or in the evaluation of indeterminate nodules detected through FNAB.

In general, the choice of imaging modality depends on various factors, including the patient's clinical history, physical exam findings, and characteristics of the nodule(s). A physician can decide the most suitable options based on the individual's specific situation [15,16].

1.5 Diagnosis of Thyroid Nodules using Ultrasound Imaging

Ultrasound is the most commonly used imaging modality and the first-line diagnostic tool for evaluating thyroid nodules due to its non-invasive nature, lack of ionizing radiation, and ability to provide detailed images of the thyroid gland and surrounding structures.

During a thyroid ultrasound the images generated allow clinicians to visualize the number, size, and characteristics of thyroid nodules, including aspects such as echogenicity, margin characteristics, presence of calcifications, and whether the nodule is solid or cystic.

Importantly, ultrasound can also assess the blood flow within the nodule and surrounding thyroid tissue using color Doppler flow imaging. Increased blood flow can sometimes be associated with malignancy.

To stratify the risk of malignancy in thyroid nodules and help decide the necessity of further steps like fine needle aspiration biopsy (FNA) or scintigraphy, various classification systems have been developed, among which the Thyroid Imaging Reporting and Data System (TIRADS) is commonly used.

1.5.1 Thyroid Imaging Reporting and Data System (TIRADS)

TIRADS is a risk stratification system developed to assess the malignancy risk of thyroid nodules detected on ultrasound, and to guide decision-making about the need for a FNAB.

To elaborate on an example, according to Kwak et al. [17] their system uses suspicious ultrasound features to determine the TIRADS category of a nodule: solid component, hypoechogenicity, marked hypoechogenicity, microlobulated or irregular margins, microcalcifications, and taller than-wide shape. The TIRADS category in turn corresponds to a specific risk of malignancy. The categories include TIRADS category 3 (no suspicious US features), 4a (one suspicious US feature), 4b (two suspicious US features), 4c (three or four suspicious US features), and 5 (five suspicious US features). The higher the category, the higher the recommendation of an FNAB.

While TIRADS has been widely adopted and has significantly improved the management of thyroid nodules, it should be noted that there are several versions of TIRADS, including those developed by Kwak et al. [17], the American College of Radiology (ACR) the American Thyroid Association (ATA) [18], and the EU-TIRADS [19], among others. Each of these systems has its own criteria and scoring system, reflecting the ongoing research and debate in this area. As such, clinical judgement and patient preferences should also be taken into account when deciding on the best approach to a thyroid nodule. Due to the presence of such variability that arises during the use of these reporting systems [17-19] a large amount of subjectivity and inter- and intraobserver variability is seen in the thyroid nodule diagnostic process using US images. Due to which, the decision making process tends to suffer from inconsistencies leading to the lack of quality healthcare delivery to the patient. This poses the need for a

computer aided decision support system that can help introduce objectivity in the decision making process as well as reduce the observer variability.

1.6 Technical State-of-the-art

This section gives an overview of some of the technologies that are state-of-the-art (SOTA) used in the detection, risk-stratification and classification of diseases using medical images. It highlights concepts such as texture analysis, Machine Learning (ML), Deep Learning (DL) and shape-based features. It must be noted that the techniques and examples used in this section can most of the time be applied to medical images irrespective of the modality. The selection of which technique needs to be used is dependent on the problem statement. Though the field of image processing and computing is vast, the concepts given in this section are directed more towards their use with respect to thyroid nodule/cancer diagnosis through computer aided methods.

1.6.1 Texture Analysis

Image texture analysis is a critical step in computer-aided diagnosis systems. Texture features reflect the spatial distribution of pixel intensities particularly in medical images, potentially capturing microstructural information that may be difficult to perceive by the human eye. Additionally, these features act as inputs for algorithms aimed at aiding in pattern recognition, clustering, classification and segmentation as part of larger problem statements. Algorithms like Grey Level Co-Occurrence Matrix (GLCM) [20], Grey Level Run Length Matrix (GLRLM) [21], and Local Binary Patterns (LBP) [22] have been utilized in several studies for texture feature extraction. Texture analysis is a fundamental process in the evaluation of thyroid nodules in ultrasound images. It refers to a variety of mathematical techniques that characterize the pixel intensity distribution in an image region. These texture features can capture information about the inherent microstructure and patterns of the tissues, potentially revealing

insights about the nature of the thyroid nodules. The process of texture analysis in ultrasound images involves several steps and can be achieved through a combination of image processing and signal processing. The goal is to quantify specific characteristics or patterns in the pixel intensity distribution that may be associated with the target or diseased region of interest's properties. In this context, a region of interest could refer to either a nodule, tumor, region or the thyroid gland itself.

Several categories of texture features are commonly used:

1. Statistical Features: These are derived from the statistical operations and distribution of pixel intensities. Examples include mean, variance, skewness, kurtosis, and entropy. Histogram-based features also fall into this category.

2. Model-Based Features: These are based on mathematical models that represent the texture. For instance, autoregressive models have been used to represent the spatial dependencies of pixel intensities.

3. Transform-Based Features: These involve transforming the image into a different space and extracting features from the transformed image. For example, the derivation of signal information from images to further analyse them in the frequency domain as opposed to a standard time domain. Examples include Fourier Transform, Wavelet Transform, and Gabor Filters.

4. Structural Features: These involve defining specific spatial patterns (or "structures") and determining how often they appear in the image [23].

One of the most popular methods for texture analysis is the Gray-Level Co-Occurrence Matrix (GLCM) [20], which considers the spatial relationship of pixels. It calculates how

often pairs of pixels with specific values and spatial relationships occur in an image, creating a matrix from which several texture measures can be extracted, such as contrast, correlation, energy, and homogeneity. Another method is the Run-Length Matrix, which counts the length of consecutive pixels that have the same gray level values [24].

1.6.2 Machine Learning

Machine learning (ML) represents a significant advance in the classification of different thyroid textures in ultrasound images. It uses mathematical algorithms that improve automatically through experience. Particularly in medical imaging, ML methods provide a valuable tool to aid in distinguishing between different thyroid tissue types based on various features previously extracted. The initial step in this process involves feature extraction from ultrasound images. Features can be texture-based, capturing the patterns and interrelationships of pixel intensities within the image, or they can be shape-based, encoding the geometrical properties of the nodules. These features are instrumental in capturing the underlying micro-structure of thyroid tissue, which can vary between benign and malignant nodules. Once these features are extracted, they are used as inputs to train machine learning models, which can learn the patterns in the feature space that correlate with different types of thyroid tissues. There are several ML techniques available. The ones given below are just a few that were employed in this work and are meant to provide a very basic understanding.

1. Support Vector Machines (SVM): SVM is a popular choice for the classification of thyroid nodules due to its robustness in high-dimensional feature space and its ability to handle non-linear classification problems using kernel functions. It aims to discern complex patterns within image datasets. The algorithm establishes an optimal decision boundary, also known as a hyperplane, within a multidimensional space of extracted

features. This hyperplane delineates distinct classes, such as healthy tissue and pathological anomalies like tumors.

2. Random Forest Classifier (RFC): Random Forest Classifiers aggregate the predictions of multiple decision trees through bagging and feature randomness, offering excellent accuracy and resistance to overfitting. This approach can be understood by looking at it as collecting diagnoses from multiple experts and then averaging them to find the best and fairest possible outcome.

3. k-Nearest Neighbors (k-NN): The k-NN algorithm classifies images based on the class of their nearest neighbors in the feature space. The "neighbors" refer to the existing labeled data that acts as a guide for the algorithm to find similar patterns in unlabelled data. It is a simple yet effective method, especially when the distribution of the data is unknown.

4. Artificial Neural Networks (ANN): ANNs consist of interconnected neurons that transform the input features through several layers to perform classification. The algorithm identifies patterns it has seen before and finds similarities in new images to provide a result. Despite requiring more computational resources and being more prone to overfitting, ANNs can capture complex, non-linear relationships in the data.

The performance of these models is assessed using a variety of metrics, such as accuracy, sensitivity (SE), specificity (SP), Dice Coefficient (DC) and area under the receiver operating characteristic curve (AUC-ROC).

While these ML methods hold great potential, there are important considerations. The performance of the models largely depends on the quality of the features extracted from

the ultrasound images. Furthermore, given that these models learn from the data they are trained on, the diversity and size of the dataset, as well as the accuracy of the annotations (ground truth), play a vital role in the model's ability to generalize well to unseen data [25].

Image segmentation, another crucial step, involves the isolation of the region of interest (in this case, the thyroid nodule) from the background. This process can be conducted using several image processing techniques such as thresholding, edge detection, region growing, and more recently, advanced deep learning models like U-Net [26], ResNet-50 [27] or nnU-Net [28]. The effectiveness of segmentation directly influences the accuracy of the subsequent feature extraction and classification as previously seen in machine learning approaches.

1.6.3 Deep Learning

Deep learning, a subset of artificial intelligence, has shown remarkable potential in medical imaging, particularly in the detection and segmentation of thyroid nodules in ultrasound images. This technology utilizes multi-layered artificial neural networks that can learn directly from data, automatically extracting complex patterns that can be used to classify images or segment features of interest. Deep learning can assist physicians in detecting the presence of thyroid nodules within ultrasound images. Models such as Convolutional Neural Networks (CNNs) have been particularly effective due to their ability to process image data directly, removing the need for manual feature extraction which is required in traditional machine learning models. CNNs use convolutional, pooling, and fully connected layers to analyze local and global patterns within images. When trained on large datasets of thyroid ultrasound images, these models can learn to identify the subtle characteristics that may indicate the presence of a nodule [29]. Once nodules are detected, accurate segmentation, or the process of delineating the boundary

of the nodule from the surrounding thyroid tissue, is critical for subsequent analysis and diagnosis. However, due to the heterogeneity and irregular shape of thyroid nodules, plus the presence of speckle noise in ultrasound images, this task can be challenging.

Deep learning models have shown promising results in overcoming these challenges. Specifically, U-Net, a type of CNN, has become a popular choice for medical image segmentation tasks. U-Net's architecture, consisting of a contracting path to capture context and a symmetric expanding path for precise localization, allows it to make accurate segmentations even with a limited number of training images [26]. Other deep learning models used in segmentation tasks include DeepLabv3 and Mask R-CNN [30]. Further variations on such algorithms also exist such as [28] and continue to improve the field of segmentation.

It's worth noting that the performance of these deep learning models depends significantly on the quality and diversity of the training data. Models need to be trained on a large, diverse set of thyroid ultrasound images to ensure they generalize well to new, unseen images. Additionally, because these models learn from the annotations provided in the training data, it's crucial that the ground truth annotations are accurate and reliable. Deep learning holds significant promise for the detection and segmentation particularly of thyroid nodules in ultrasound images. It can potentially increase the efficiency, consistency, and accuracy of these tasks, aiding in the probable early detection and appropriate management of thyroid conditions. However, as with any technology, careful validation and interpretation of results are necessary.

1.6.4 Shape-based Features

Geometric and morphological feature (G-M) extraction is one of the potential subsequent steps after the segmentation of a region of interest. Such features help in understanding the size, shape, and structural arrangement of several anomalies found in medical images, such as nodules, masses or tumors. Features like area, perimeter, compactness, elongation, and other shape-based descriptors can provide vital clues about the nature of the nodule.

Geometric and morphological feature extraction plays a pivotal role in the analysis and classification of thyroid nodules in ultrasound images. These features provide insights into the physical shape and structure of the nodules, and when used in combination with other types of features, such as texture-based or intensity-based features, they can significantly improve the performance of classification algorithms [31]. These features can be calculated directly once the nodule region has been segmented from the ultrasound image. For instance, a study by Weinmann et al. found that geometric features such as solidity and extent, which measure the compactness of the nodule, could distinguish between benign and malignant nodules with high accuracy [32].

Morphological features go a step further, capturing the spatial arrangement and configuration of the tissues within the nodule. Morphological features often require more sophisticated image processing techniques to calculate. For instance, fractal analysis can be used to calculate the fractal dimension of the nodule, a measure of its complexity. Shape descriptors, such as circularity, eccentricity, and convexity, can provide further clues about the nature of the nodule. In addition, the morphology of the nodule's boundary can be examined. Features like roughness, irregularity, or lobulation of the boundary have been shown in several studies to correlate with the malignancy of the nodule.

Once these geometric and morphological features have been extracted, they can be used as input to machine learning algorithms for the classification of the nodules. Algorithms such as SVM, Decision Trees (DTs), RFCs, and/or k-NNs have been used to classify nodules based on these features.

However, it is important to note that while these features can be informative, they are not entirely infallible. The performance of these techniques can be influenced by several factors, including the quality of the ultrasound image, the accuracy of the segmentation, and the variability in the appearance of thyroid nodules. Furthermore, these techniques are most effective when used in combination with other forms of analysis, such as texture analysis and/or clinical information. Therefore, while geometric and morphological feature extraction provides valuable tools for the classification of thyroid nodules, it should be considered as part of a larger, multifaceted approach to nodule analysis and diagnosis.

1.6.5 Application of Discussed State-of-the-Art Technologies in Thyroid Ultrasound Images

This section provides an overview of the previously highlighted techniques and their applications in literature.

Consider a cascading approach to the diagnostic process as shown below.

1. Moving from the outer surrounding tissue, to the detection of the thyroid gland for the localization of the region of interest (thyroid gland)
2. Segmentation of thyroid nodules
3. The region estimation in thyroid nodules
4. Subsequent classification of nodules

The following sections briefly introduce each of the aforementioned aspects from the cascading approach and then proceed to highlight relevant work done by researchers in that particular discipline.

Starting with the detection of the thyroid gland in US images, from a research standpoint, numerous studies have examined various approaches to segmenting the thyroid in individual 2D US images. Zhao et al. [33] proposed multiple methods, such as edge detection, threshold value method, region splitting and merging, watershed segmentation, active contour, graph theory, US image segmentation based on N-cut, and segmentation based on improved normalized cut.

Kaur and Jindal [34] explored the segmentation of the thyroid from 2D US and scintigraphy images using active contour without edges, localized region-based active contour, and distance regularized level set. In another instance, a polynomial SVM was employed [35] to segment the thyroid gland in US images. A study [36] proposed using a local region-based active contour to segment and compute the area of the segmented thyroid in a 2D US image. Mylona et al. [37, 38] implemented a similar region-based active contour for medical image segmentation, encoding the local geometry information to control the contour evolution. Keramidas et al. [39] proposed thyroid segmentation in US images using a novel boundary detection method and local binary patterns for texture analysis. Other researchers used level-set active-contours models for thyroid segmentation in US images [40, 41], focusing on variable background active contour and joint echogenicity texture. Garg and Jindal later employed a feed-forward neural network [42] for thyroid gland segmentation from US images. More recently, Narayan et al. [43] used speckle-related pixels and imaging artifacts to perform multi-organ segmentation in thyroid US images.

Furthermore, several studies have aimed to segment a full 3D thyroid image. Kollorz et al. [44] put forward a semi-automated approach for volumetric quantification of the thyroid gland using geodesic active contour. Chang et al. [45] proposed using a radial basis function (RBF) neural network to segment the blocks of the thyroid gland. Iakovidis et al. [46] utilized the fusion of fuzzy statistical distributions in order to determine how tissue patterns occur in the thyroid gland in ultrasound images. Finally, Osman [47] conducted a complete segmentation and analysis of 3D thyroid images using a method that involved thresholding voxel intensities and connecting similar voxels to predict the segmenting regions.

In addition to this Poudel et al. [48, 49] utilized Active Contours Without Edges (ACWE) models to segment thyroid glands in 2D thyroid US images. They then employed these segmented 2D images to create 3D representations, a process aimed at achieving 3D thyroid image segmentation [50].

Another notable contribution to the field was by Illanes et al. [51] They proposed a unique technique for image texture feature extraction in US images, predominantly based on parametric modeling. The principal idea of this method was to treat the texture as data derived from a dynamic process, and identify the various dynamics embedded in the texture. This enabled them to employ mathematical operations among these dynamics to yield potential texture features. In executing this approach, a signal representation of the image was initially generated, where space functioned as the independent variable. This signal was subsequently broken down into distinct frequency bands through Wavelet Transformation. In the final stage, an Autoregressive (AR) parametric model was applied to these decomposed signals, yielding spectral characteristics that were harnessed for the computation of features.

In their research, Chang et al. [52] utilized the Decision Trees (DT) model for segmenting thyroid nodules in 2D thyroid ultrasound images. Their methodology comprised two stages: image preprocessing and image segmentation. During preprocessing, histogram equalization was first employed to enhance the contrast between nodules and their background, followed by the determination of potential nodular areas through horizontal and vertical projections. Then, 41 features were extracted using eleven feature extraction techniques, including co-occurrence matrix [53], statistical feature matrix [54], gray level run-length matrix [55], and more. For image segmentation, decision trees were used to construct a classifier which classified image blocks into either nodular or background regions. The accuracy of the segmentation results was 97.5% for a test dataset of six 2D thyroid ultrasound images.

In parallel, Keramidas et al. [56] proposed a Thyroid Nodule Detection system for analyzing 2D thyroid ultrasound images and videos. This system's detection process consisted of five steps, starting from data preprocessing, definition of the region of interest, feature extraction classification and ending at post-processing. The preprocessing of data essentially refers to the cleaning and preparation of data before any mathematical or annotation activities can be carried out. In this case the authors used image normalization that converts the image and pixels into the same range and format. The definition of the region of interest involved the determination of the ground truth. Textural features were extracted and techniques like SVM [57] and k-NN were used for the classification stage. Misclassified pixels and regions in the output binary images of the classifier were reduced using a majority voting decision criterion [58] and referred to as post-processing. The segmentation accuracy of thyroid nodules achieved by SVM was 91.3%, tested on a dataset of 118 2D thyroid nodule ultrasound images.

Moving towards the segmentation of thyroid nodules and their classification, it can be seen that there exists several techniques to achieve the same goal. For instance, deep neural networks have excelled in tasks like semantic segmentation and object recognition and detection. Ma et al. [59] employed a deep convolutional neural network to segment thyroid nodules in 2D thyroid images. Their model included fifteen convolutional layers and two max pooling layers, and was trained and validated using tenfold cross validation on a dataset of 22123 ultrasound images. The trained CNN model yielded a mean overlap value of 86.83%. Distinctively, Ying et al. [60] proposed a phased CNN model called the cascaded convolutional neural network (CCNN) to segment thyroid nodules. This model differed from the end-to-end CNN model described by Ma et al. [59] and was composed of three phases. The CCNN model involved manual sign recognition and boundary adjustment to make artificial marks around the nodules. The mean overlap value of the segmentation result on the testing set was 87.00% [59].

Kumar et al. proposed a novel multi-output convolutional neural network algorithm with dilated convolutional layers to segment thyroid nodules, cystic components inside the nodules, and normal thyroid gland from clinical ultrasound B-mode scans. Through their prospective study, they eliminate the need for a seed in the segmentation process, thus automatically detecting and segmenting the thyroid nodules and cystic components with a dice coefficient of 0.76. [61]. Similarly, Nugroho et al. attempted to classify the internal contents of a thyroid nodule based on textural features such as histogram statistic, gray level co-occurrence matrix (GLCM) and gray level run length matrices (GLRLM). Once features were extracted, the authors used a multi-layer perceptron (MLP) to obtain a classification of the solid and cystic components with an accuracy of 90.28%, the sensitivity of 87.80%, specificity of 93.55% and precision of 94.74% [62]. Zulfanahri and team proposed a method to extract shape-based features

contributing to the classification of thyroid nodules in Ultrasound images. They successfully extracted 7 geometric features and 14 moment features. The selected features were classified using a Support Vector Machine and achieved 91.52%, 91.80% and 91.35% as metrics for accuracy, sensitivity and specificity respectively [63].

1.7 Motivation and Contribution

1.7.1 Motivation

The relevant literature presented above has its own set of limitations. From a clinical standpoint the current diagnostic process for thyroid nodules using US images is subjective and consists of a high degree of inter- and intraobserver variability due to the fact that it depends on the experience level of physicians. This is a cause for concern as these factors potentially act as risks to the patient's health and their improvement. This is because it makes the process biased, longer and sometimes conflicting that does not help a patient in receiving optimal care or in improving their quality of life. From a technical point-of-view, in terms of thyroid gland detection, current literature deals with approaches that do not differentiate between the textures of the thyroid gland and the surrounding structures adequately due to the presence of speckle noise in US images. The delineation of the thyroid gland helps the assessing physician focus on the area to be examined with a higher degree of objectivity. Once focused, the next step is to easily identify the nodules present in the thyroid gland. Literature consists of several deep learning models that are capable of doing so. However, to aid in the diagnostic workflow, which of these algorithms is best suited for the task still needs to be highlighted. As we move through the diagnostic process for thyroid nodules using ultrasound images, solid and cystic region estimation plays a pivotal part that improves subjectivity and also helps physicians better quantify and estimate these regions. An example for such a need would be for improved radioiodine dose calculation. The current literature does not address this as a clinically relevant parameter when it ought

to do so. With TIRADS being a crucial part of the diagnostic process, particularly when it comes to visual characteristics of the nodules. Current approaches do not take into consideration that these characteristics are what physicians look at as well during the classification process. Overall, current approaches fall short when it comes to taking a systematic approach that considers the entire diagnostic process as a whole in a way that addresses the needs and understanding of the physician.

To summarize, the current state of the art (i) does not differentiate between textures of the thyroid gland and surrounding structures, (ii) rarely states the selection of a proper deep learning model for the segmentation and localization of thyroid nodules in US images, (iii) provides no methods for the estimation and quantification of solid and cystic regions within the nodule to facilitate objectivity and aid in dose calculation for radioiodine therapy, and (iv) does not address the lack of approaches that employ shape-based features that ought to be considered in accordance with the visual characteristics currently used by physicians through TIRADS. Addressing these shortcomings through a systematic approach for gland detection, nodule localization, nodule region estimation and overall nodule classification would result in the reduction of subjectivity and inter- and intraobserver variability during the diagnostic process.

1.7.2 Contribution

The work presented herein contributes to the domain by investigating and developing techniques using image processing, texture analyses, machine learning and deep learning for the detection, region quantification and classification of thyroid nodules. The outcomes of which would support physicians by providing objective and quantitative assessment of thyroid nodules and hence reduce subjectivity and inter- and intra-observer variability. Additionally, the work presented will also take into consideration the features looked at by the physicians during the current diagnostic

process for thyroid nodules in US images so as to provide a better understanding. An overview of the proposed approach and techniques utilized while depicting the contributions of this work to the domain are given in Figure 3. Figure 3 also provides a reasoning for each publication, hence attempting to justify the systematic nature of the approach. The proposed approach begins with the detection of the thyroid gland using autoregressive features in conjunction with machine learning for the analysis and differentiation of textures from the surrounding structures. This was done to focus the attention of the physician on the region to be examined. The next step involves a selection of an appropriate deep learning algorithm for the segmentation and localization of thyroid nodules. For this, a comparison of four well known deep learning algorithms for segmentation were trained and their performance was compared to determine the one that best suits the data and problem statement. The third step involves the use of texture analysis and machine learning in the estimation of solid and cystic regions within the thyroid nodule. This was done to provide the physicians with a quantification of these regions that would aid in objective dose calculation for radioiodine therapy. The final step utilizes shape-based features (geometry and morphology) together with machine learning approaches to classify thyroid nodules as either benign or malignant. This was done in an attempt to capture the visual characteristics seen by the physician during the use of TIRADS and overall diagnostic process.

For the development of any form of computer aided diagnostic system, data plays a key role. Apart from using an open-source dataset [64], this work additionally required more and varied forms of data. This was needed to (i) gain a better understanding of the current methodologies and techniques used in US based thyroid nodule assessment, (ii) get involved in the clinical workflow and practices, (iii) establish a data collection protocol that could be used in future studies and (iv) collect data that could help in the

analysis and development of the solutions presented in this work. As an additional contribution to the work a data collection strategy was developed by submitting an ethics proposal to the ethics commission (RAD362-16/19) at the Otto-von-Guericke University Hospital, Magdeburg. Once ethical clearance was granted, the data collection was carried out twice a week for a period of 8 months. The data was collected in the form of B-mode US videos together with an electromagnetic tracking system using a 9L probe and LOGIQ S8 system from GE (Wisconsin, USA). A total of 47 patients with 78 nodules were found from which 2290 2D-US images containing nodules were extracted [65]. Three experienced physicians annotated these cases for the generation of the ground truth in terms of nodules, regions and classification. Each physician first annotated these images independently and a consensus was achieved through discussions for cases with conflicts. These images have contributed throughout this work either in the form data to be used for analyses and development or as a means to better understand the behavior of thyroid nodules in US images. For development, this data was coupled with the earlier mentioned approaches to detect the thyroid gland, localize thyroid nodules, estimate nodule regions and provide an overall classification of nodules using visual shape-based features to support decision making.

The upcoming sections of this thesis are divided as follows. Chapter 2 highlights the hypotheses for each of the publications and briefly describes the methodology designed to support them. Chapter 3 presents the results of the selected approaches according to the publications. Chapter 4 discusses these results together with what is found in literature while concluding and providing the future scope of the work and summary of the report.

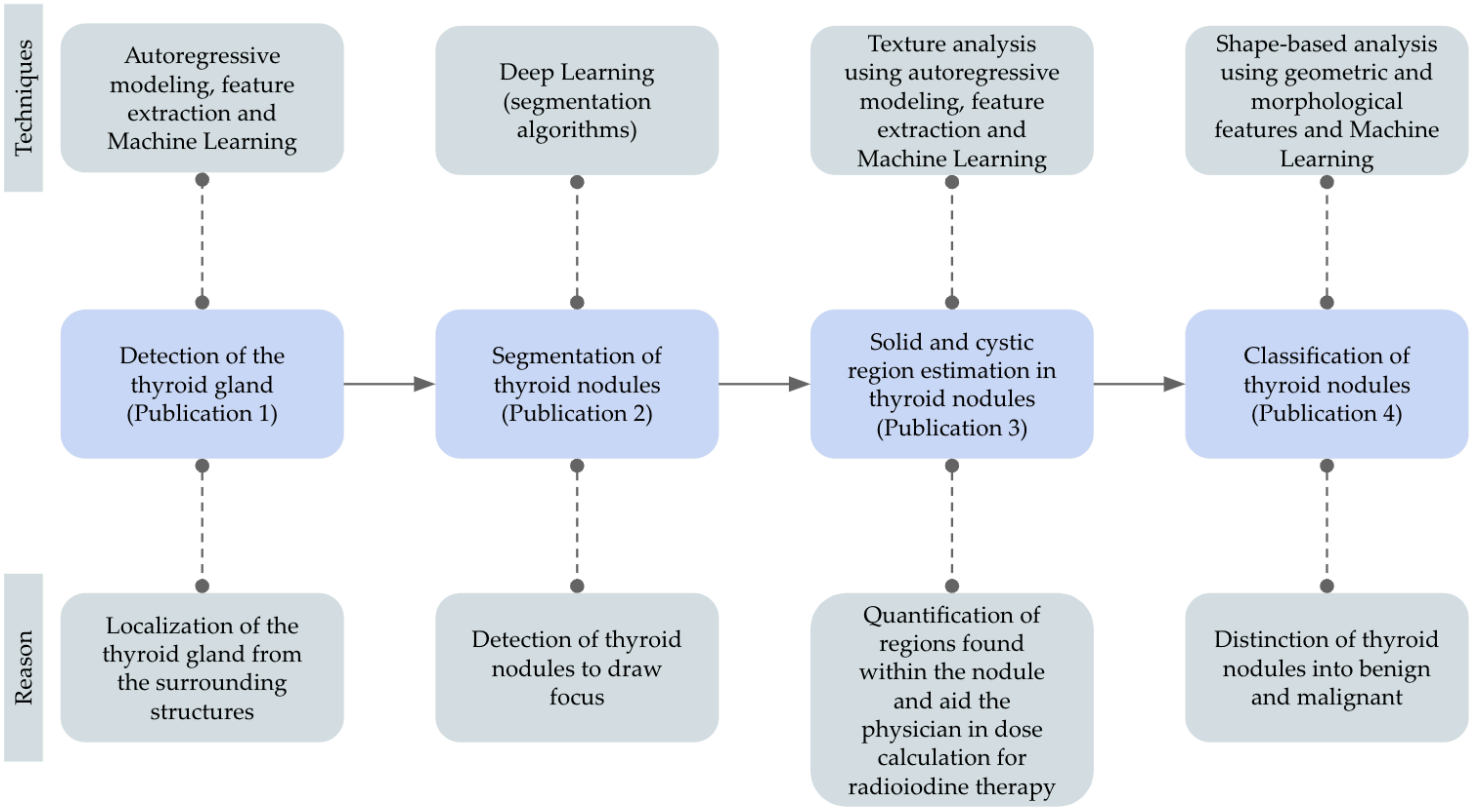


Figure 3: Flow diagram of proposed approach. The boxes in the middle (in blue) highlight the general path followed for the work according to the publications. The boxes (in green) at the top represent the specific feature extraction, machine and deep learning techniques used in each publication while those at the bottom provide a reasoning for the work in contribution to the systematic approach

Chapter 2: Hypotheses

Based on the background and motivation, the following hypotheses were formulated for each of the publications included in this report.

H1: The combination of autoregressive features and Machine Learning approaches can help distinguish between textures of the thyroid and the surrounding structures thus contributing to more accurate and objective localization of the thyroid gland.

H2: There is a significant difference in the performance of various Deep Learning algorithms in the semantic segmentation of ultrasound images of thyroid nodules, with one or more of these algorithms demonstrating superior efficacy in accurately segmenting and characterizing these nodules.

H3: The combination of autoregressive modelling and Machine Learning techniques will yield an effective method for estimating the region of thyroid nodules, resulting in quantifiable evidence to support decision-making.

H4: Machine Learning-aided evaluation of geometric and morphological features can effectively classify thyroid nodules, thereby enhancing the decision-making process of physicians by considering visual characteristics of the nodules.

2.1 Methodology

2.1.1 Detection of the Thyroid Gland using Features Extracted from US images and Classification using Machine Learning for the Localization of the Thyroid Gland (H1-Publication 1)

By using textural analysis for feature extraction in conjunction with machine learning models, we aim to detect the thyroid gland so as to localize and focus the area that would be analyzed during the diagnostic process. To achieve this, two datasets were used. Dataset 1 consisted of 675 and Dataset 2 consisted of 3,370 images containing the thyroid gland. These images were annotated by experts and then split into non-overlapping texture patches with a size of 20x20 pixels as seen in Figure 4. A total of 90,816 and 1,791,397 such texture patches were extracted from the original thyroid gland images for Dataset 1 and Dataset 2. The texture patches were then converted into signals using four different types of transformation patterns. The transformation from matrix to signal was done using ZigZag and Spiral transformation and also using their 90° rotated patch version as depicted in Figure 5. This was done to capture the texture dynamics found in the patches. The transformed signals were then decomposed into 4 narrow band signals (low, mid, high and total frequencies) using Continuous Wavelet Transform (CWT). These frequency bands were then modeled using Autoregressive modeling and 30 significant features were extracted. The extracted features were then fed into three different machine learning models to obtain texture classifications (SVM, ANN and RFC) for Datasets 1 and 2. The performance of these models was also compared to each other and the state of the art. A post processing step was also employed in the form of the largest connected component analysis to compensate for some over-classified texture patches.

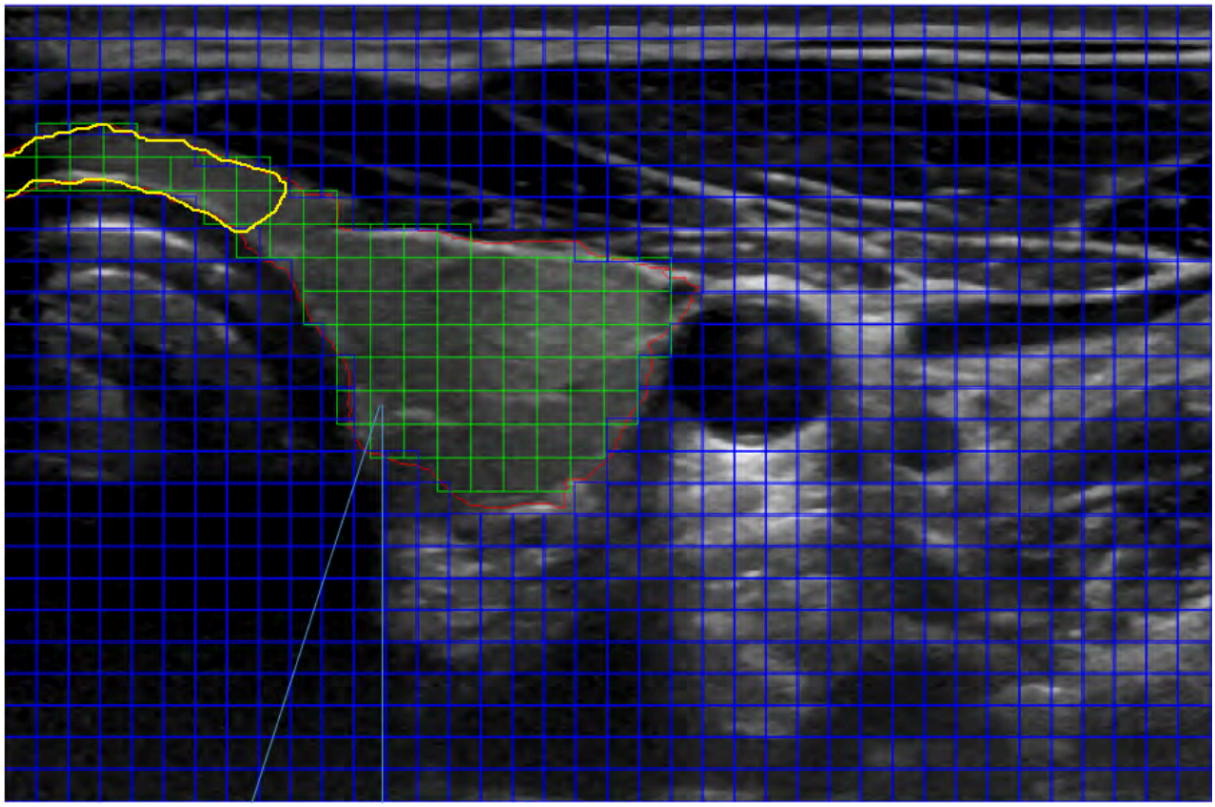
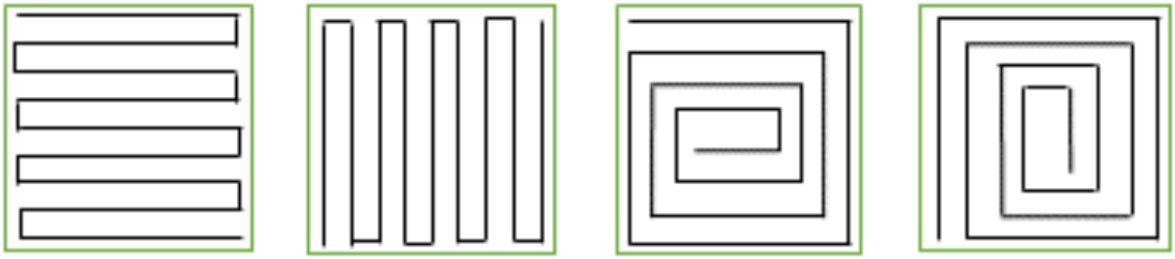


Figure 4: Extraction of 20x20 texture patches from the thyroid gland and surrounding structures collected for the feature extraction



(i) (ii) (iii) (iv)

Figure 5: Conversion of the extracted texture patches using four different transformations of matrix to a signal. (i) ZigZag, (ii)ZigZag rotated by 90°, (iii) Spiral and (iv) Spiral rotated by 90°

2.1.2 A Comparison of Deep Learning Architectures for the Detection of Thyroid Nodules in US Images (H2 - Publication 2)

The use of DL is aimed at detecting the nodule within the thyroid gland so as to focus the area of investigation. DL models have the ability to consider features that are not visible to the naked eye but still quantitatively relevant. The use of deep learning for the aid in an objective detection of thyroid nodules that can be then used for further analysis. A total of 1011 ultrasound images containing thyroid nodules were used in this study. Data preprocessing techniques were used to remove background noise, identify contours and resize images to bring them to the same scale. Following this data augmentation techniques such as horizontal and vertical shifts, rotation at angles in the range of 5 to 15 degrees, different affine transformations, and gray value variations were employed. Random elastic transformations for the purpose of generating more generalized results were also used. The transformed data was then fed into four SOTA deep learning algorithms U-Net [26] (architecture depicted in Figure 6), ResUNet [66] (architecture depicted in Figure 8), SUMNet [67] (architecture depicted in Figure 7), and Attention U-Net [68] (architecture depicted in Figure 9). The results from each of the algorithms were compared in terms of Accuracy, Dice Coefficient (DC) and Intersection over Union (IoU).

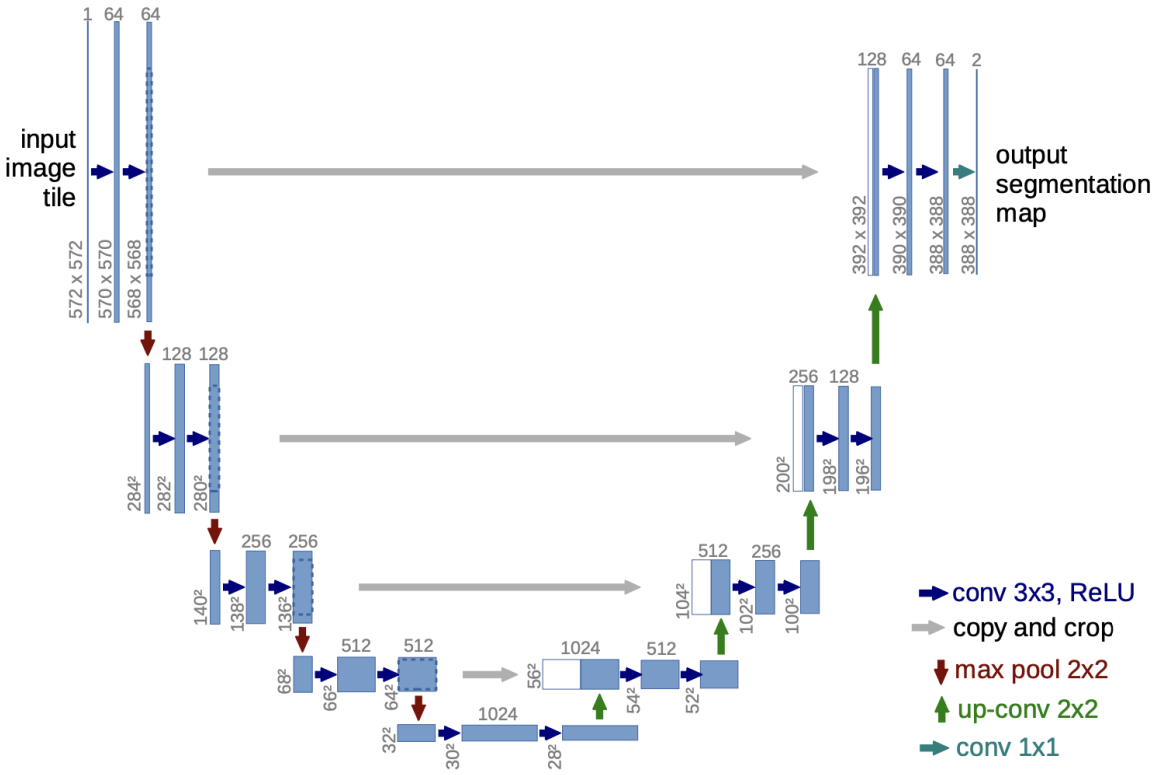


Figure 6: U-Net architecture as proposed by Ronnenberger et al. [26] that is used for the segmentation of structures in biomedical images

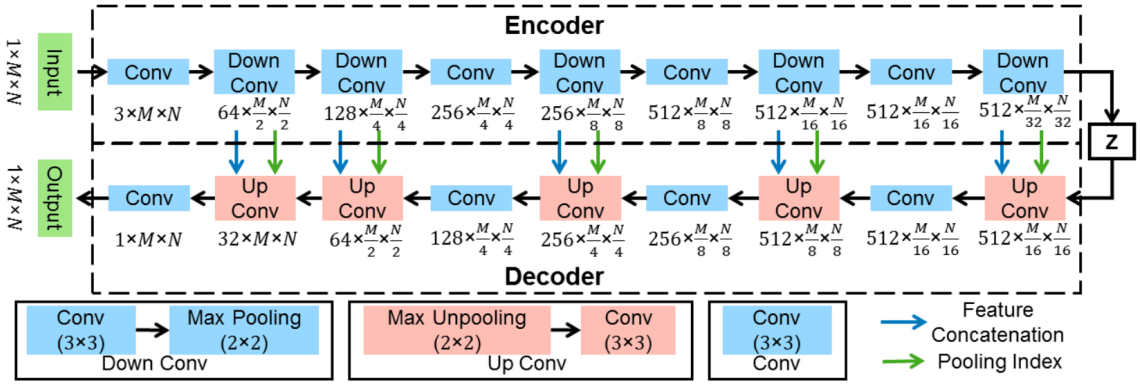


Figure 7: SUMNet architecture as proposed by Nandamuri et al. [67] developed specifically for the segmentation of structures in ultrasound images

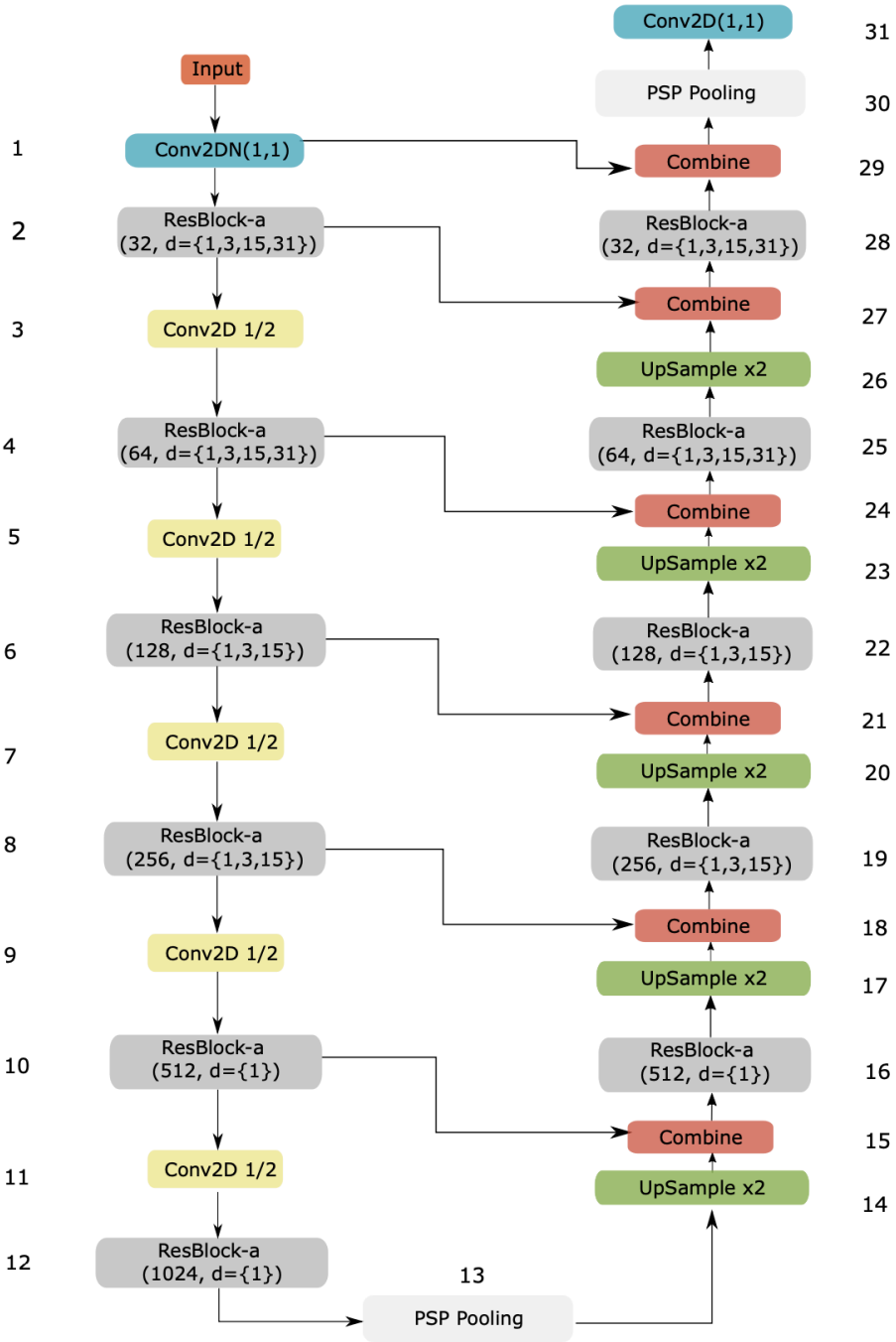


Figure 8: ResUnet architecture as proposed by Diakogiannis et al. [66] used for the segmentation of images and uses a residual layer together with the architecture of the U-Net

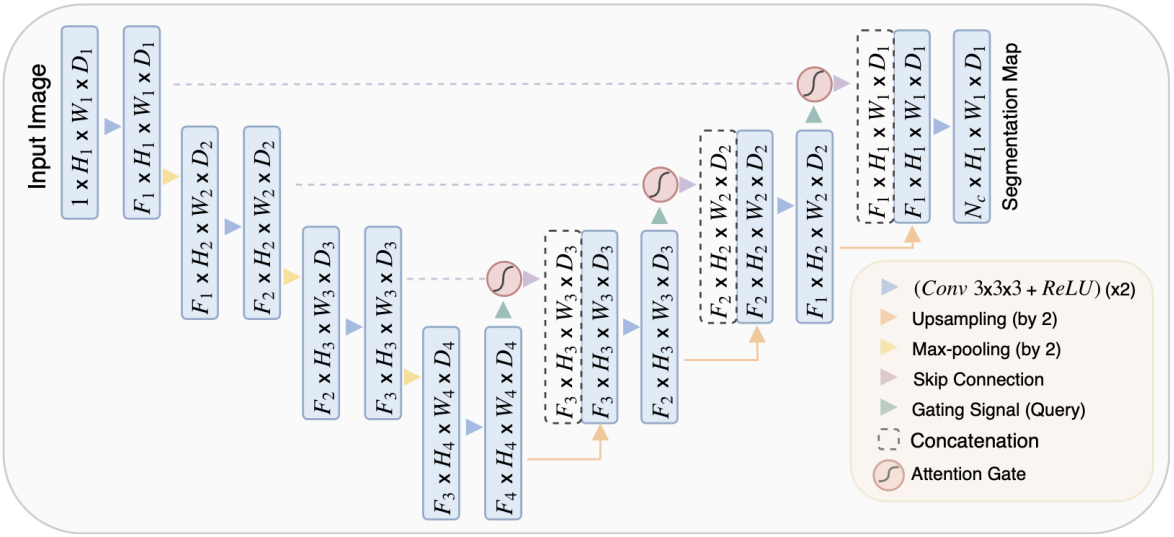


Figure 9: Attention U-Net architecture as proposed by Oktay et al. [67] that helps in segmentation tasks by combining a U-Net architecture with attention gates

2.1.3 Solid and Cystic Region Estimation in Thyroid Nodules using US Images to Facilitate Objective Decision Making Among Physicians through Quantifiable Evidence (H3 - Publication 3)

Differentiating between cystic and solid regions in a nodule aids in the risk-stratification. Multiple studies have shown that the portion and percentages of cystic and solid components in a thyroid nodule can be indicative whether a nodule is benign or malignant. Thyroid nodules exhibiting a larger cystic portion are considered to be benign whereas those nodules that are predominantly solid have a higher risk of being malignant. Therefore, it is helpful for the physician to know these characteristics for a newly detected nodule and knowing their percentage is aimed at better risk-stratification. Additionally, the determination of the exact percentage of solid and cystic regions would aid the physician in better calculation of doses for radioiodine therapy. A total of 1019 US images were used in this study. Each of these images consisted of nodules containing solid and cystic regions. The regions were annotated

based on the expert input of 4 physicians. Texture patches of size 20x20 pixels were extracted from each of the solid and cystic regions to create a texture database. The database consisted of a total of 10190 solid texture patches and 8980 cystic patches. Each of the texture patches were converted into signals by transformation from matrix to signal using ZigZag and Spiral transformation and also using their 90° rotated patch version. Each of the signals were decomposed into 3 frequency bands using CWT decomposition. Low frequency bands were selected between the ranges of 8-24 Hz. The mid frequencies were between 25-33Hz while the high frequencies were between 34-46 Hz. A total of 36 features were then extracted using AR modelling and used together with machine learning algorithms (ANN, k-NN and RFC) to classify the textures of the regions. The results of the machine learning algorithms were compared and the best selected. Furthermore, the performance of the machine learning algorithms with the AR features were compared to their performance using Bispectrum features on the same textural database. Once the classification was done, the estimated regions of the nodule were quantified as percentages.

2.1.4 Classification of Thyroid Nodules in US Images using Shape based Features and Machine Learning to Support Clinical Decisions with Currently Considered Visual Characteristics (H4 - Publication 4)

Geometric and morphological feature extraction that takes into consideration features that are closely related to the visual shape-based (TIRADS) features currently used by physicians. This provides them with additional information and mathematical evidence to support their current TIRADS-based classification with an extra layer of objectivity. In this work, we mainly focus on shape-based geometric and morphological feature extraction for the classification of thyroid nodules as either benign or malignant. Examining physicians use visual and textural characteristics to classify a nodule. Geometric and morphological features represent the visual aspect. By using

shape-based features such as geometry and morphology we are able to emulate the visual characteristics of the thyroid nodules that are looked at by the physicians. Coupling these features with Machine Learning gives us the appropriate class of a thyroid nodule (benign or malignant). An open-access database consisting of 99 cases featuring thyroid nodules in US images was used. Data augmentation techniques were used to balance and augment the data to obtain a total of 3188 images. A total of 27 features (19 geometric and 8 morphological) were extracted from the images. The features were selected based on the expert input from experienced physicians and further supported by testing the performance of the features using machine learning. This led to the selection of 11 significant features which were then further assessed by comparing their performance in classifying nodules as either benign or malignant using machine learning (RFC).

Chapter 3: Results

3.1 Results in relation to Publication 1

All three classification methods—SVM, ANN, and RFC—demonstrate good, if not equal, performance in classifying thyroid texture patches. SVM exhibits the lowest accuracy among the three, achieving DCs of 0.895 and 0.887 in Datasets 1 and 2, respectively. In contrast, ANN showed the highest accuracy, with DSC of 0.930 and 0.894 for Datasets 1 and 2, respectively. RFC's accuracy is comparable to that of ANN, with DSC of 0.925 and 0.891 for Datasets 1 and 2, respectively. Additionally, performance matrices in terms of DC, SE, and SP from four distinct methodologies found in literature—Echogenicity-based Quantization (EBQ), Joint Classification Regression (JCR), Radial Basis Function (RBF), and Feedforward Neural Network (FNN)—despite these being tested on different datasets than those used in the study have been provided. Compared to other methods in literature such as Active Contours Without Edges (ACWE), Graph Cut (GC), Pixel based classifier (PBC) the features extracted coupled with machine learning methods such as SVM, ANN, and RFC consistently achieve superior DSC. Hereby validating the robustness of our feature extraction process for thyroid texture classification. Furthermore, the results for each of the machine learning algorithms are also represented visually for Datasets 1 and 2. The details of this can be found in Publication 1 in the appendix.

3.2 Results in relation to Publication 2

This segment details a comparative study of four Deep Learning segmentation algorithms: U-Net, SUMNet, ResUNet, and Attention U-Net. Each of these algorithms underwent training using the same dataset, and their results were evaluated using performance metrics such as accuracy, the Dice coefficient, and Intersection over Union (IoU), all in the context of segmenting thyroid nodules in ultrasound images. The

study's findings indicated that the U-Net architecture performed the segmentation of the thyroid nodules with 79.7%, 0.766 and 0.635 accuracy, DC and IoU respectively. The SUMNet architecture presented metrics along the lines of 83.1% accuracy, 0.795 DC and 0.672 IoU. Attention UNet performed fairly well with an accuracy of 85.4%, 0.821 DC and 0.720 IoU. But, the ResUNet architecture surpassed the others in performance, achieving an accuracy of 89.2%, a Dice coefficient of 0.857, and an IoU of 0.767. An overview of the comparison of metrics is given in Table 1. From this it can be inferred that using the ResUNet architecture particularly for the given data, it can be utilized in the further development of computer aided diagnostic tools for the detection of thyroid nodules using US images. Further details can be found in Publication 2 in the appendix.

Table 1: Comparison of performance metrics obtained using four different deep learning algorithms for the segmentation of thyroid nodules in US images

Model/Metrics	Accuracy (%)	Dice Coefficient	Intersection over Union
U-Net	79.7	0.766	0.635
SUMNet	83.1	0.795	0.672
ResUnet	**89.2**	**0.821**	**0.720**
Attention U-Net	85.4	0.857	0.767

3.3 Results in relation to Publication 3

The results obtained from this work are presented as follows. First a comparison of the performance of the three machine learning algorithms using only the Autoregressive (AR) features was carried out. Secondly, the performance of the machine learning algorithms using AR features were compared against the performance of the same machine learning models using features from Bispectrum analysis. The ANN method

recorded an accuracy, sensitivity, and specificity of 83.14%, 82.00%, and 84.00%, respectively. However, utilizing the Bispectrum-extracted features, ANN displayed an accuracy, sensitivity, and specificity of 78.29%, 79.00%, and 80.00%, respectively. The k-NN algorithm demonstrated a classification accuracy for cystic and solid regions of 84.55%, with sensitivity and specificity being 83.00% and 86.00%, respectively. In comparison, when employing Bispectrum features, the k-NN algorithm exhibited an accuracy, sensitivity, and specificity of 80.12%, 79.00%, and 88.00%, respectively, as per Table 2. Among the three algorithms, the Random Forest Classifier achieved superior results, with an accuracy, sensitivity, and specificity of 90.41%, 99.00%, and 91.00%, respectively using AR features. Additionally, the Random Forest Classifier posted an accuracy, sensitivity, and specificity of 86.32%, 87.00%, and 86.00% when Bispectrum features are used. All three algorithms successfully classify cystic and solid regions with satisfactory accuracy, as seen in both tables. However, upon contrasting the AR-modeled features and the Bispectrum features, it is clear that the AR-modeled features yield higher classification accuracy, sensitivity, and specificity. Among the machine learning models, it was determined that the RFC together with the AR features was the best suited for the region estimation in thyroid nodules and was further employed during the visualization step. An overview of this comparison is given in Table 2. Additional details can be found in Publication 3 in the appendix.

Table 2: Comparison of performance metrics obtained using three different machine learning algorithms together with two types of extracted features from the same texture database for the estimation of solid and cystic textures within thyroid nodules from US images

Features	Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)
Autoregressive	ANN	83.14	82.00	84.00
	k-NN	85.55	83.00	86.00
	RFC	**90.41**	**99.00**	**91.00**
Bispectrum	ANN	78.29	79.00	80.00
	k-NN	80.12	79.00	88.00
	RFC	86.32	87.00	86.00

3.4 Results in relation to Publication 4

In this publication, 11 selected features were fed into the Random Forest Classifier (RFC). A two-step comparison was carried out. The first step juxtaposes our proposed method with state-of-the-art methods that utilize the same dataset. The second step draws parallels between our method and other shape-based feature extraction and classification approaches utilized in related research.

In the first comparison the results showed that G-M based features attain a superior accuracy of 99.33% compared to literature while using the same database, with a high specificity score of 99.25%. Regarding sensitivity, G-M features returned a score of 99.39%. These results suggest that our feature extraction method can classify benign and malignant thyroid nodules with great accuracy, alongside high true positive (sensitivity) and true negative (specificity) rates.

In the second step, the classification results garnered from our proposed G-M feature extraction method, in comparison to two literature-sourced approaches that use shape-based features for thyroid nodule classification. The two comparative studies employ features that characterize nodules based on their oval-shaped margins. These methods utilize different datasets. Here the G-M features, when applied to the RFC classifier, showcase an accuracy, sensitivity, and specificity of 99.33%, 99.39%, and 99.25%, respectively, marking significant progress over metrics found in literature. Further details are given in Publication 4 in the appendix.

Chapter 4: Discussion, Conclusion and Future Scope

Thyroid nodules are a common clinical problem. However, the determination of whether a nodule is benign or malignant remains a challenging task. Traditional methods of diagnosis such as FNAB can be invasive, and results often depend on the experience and expertise of the physician, leading to inter- and intraobserver variability. In this regard, the application of texture analysis, shape-based feature analysis, machine learning, and deep learning techniques together with ultrasound imaging of the thyroid has been a promising development.

By utilizing a signal-based rendition of the ultrasound image which was parametrically modelled to compute Autoregressive (AR) features the localization of the thyroid gland was possible. The transformation of the image into a signal holds multiple advantages over traditional image-based analysis. In particular, the extracted features from this signal-based method remain largely unaffected by issues such as speckle noise, low contrast, and low signal-to-noise ratio prevalent in thyroid ultrasound images. This provides the classifiers with the ability to accurately delineate the thyroid region, even in smaller areas (like the isthmus of the thyroid), a task that proved challenging with several state-of-the-art methods such as Active Contour Without Edges (ACWE), Graph Cut (GC), and Pixel-Based Classification (PBC). Additional comparative analyses between this method and a variety of other approaches documented in the literature proved that the selected approach outperforms traditional approaches. The comparison was conducted in two distinct ways: firstly, by gauging performance on identical datasets, and secondly, on different datasets. In all conducted tests, our method surpassed the literature-based approaches with respect to the Dice Similarity Coefficient (DSC), while demonstrating comparable SE and SP. The results we obtained exhibited a significant correlation with the ground truth data. In terms of classifier training

accuracy, the ANN was found to slightly outdo both the SVM and RFC. Moreover, the approach was completely automated, negating the need for users to actively monitor the segmentation progress. This represents a marked improvement over techniques like ACWE, which required restarting the process if the contour initialization lay outside the thyroid region. It also eliminates the need for manual intervention in Graph Cut, where users had to remove over-classified regions post-segmentation, and in PBC, where users needed to perform multiple clicks inside and outside the thyroid regions for a more accurate estimation of features for training their decision trees. By employing a robust methodology such as this, the results obtained are indicative of the fact that thyroid gland textures can be distinguished from the surrounding structures by leveraging a combination of autoregressive features and machine learning. Hence, this method is suitable to be used as a method that contributes to more accurate and objective localization of the thyroid gland and hence helps bring into focus the region to be examined by the physician (H1).

Deep learning has emerged as a highly effective approach for semantic segmentation in medical imaging, despite the challenges posed by noise in ultrasound (US) images. Particularly, the segmentation of thyroid nodules is a critical step as it serves as the preliminary stage for comprehensive classification. Our study included a comparative analysis of four distinct deep learning architectures to identify the optimal solution for this task. The results indicate that the Residual U-Net (ResUNet) architecture surpasses other examined architectures, achieving high scores across all evaluation metrics. The standout performance of ResUNet can be attributed to its incorporation of residual blocks of convolutional layers. These blocks facilitate consistent network training as network depth increases. Moreover, these residual blocks ensure a smooth flow of gradients to the initial layers through skip connections, thereby expediting network training in the initial stages. Through the comparison of the architectures presented

here we were able to closely examine the differences in their performances when given the same data. This enabled the selection of an appropriate architecture for future work towards the development of a holistic computer aided decision support system for the diagnosis of thyroid nodules (H2).

As discussed previously, AR features prove to be robust and also have the capability to work with limited amounts of data. This approach when applied to the estimation of solid and cystic regions within a thyroid nodule yielded high quality results. When compared to other approaches such as bispectral analysis. These results when combined with the appropriate machine learning algorithm, which in this case was RFC portrays results that can classify the textures present in a thyroid nodule as either solid or cystic regions. This information in turn is used to quantify the regions and thus estimate their size in a given US nodule image in terms of percentage. Furthermore, the estimation of these regions could provide physicians with quantifiable evidence to support decision making and better planning for activities such as radioiodine therapy dose calculation (H3). However, it must be noted that deep learning when used for region estimation results in a better outcome. This needs to be kept in mind while planning for the future of this work.

When it comes to shape-based (geometry and morphology) features for nodule classification, through various feature extraction methodologies employing the same dataset, as well as similar extraction methods using diverse input data, the Geometric-Morphological (G-M) approach has demonstrated superiority over existing state-of-the-art techniques. Although comparative studies present high accuracies, sensitivities, and specificities, the features they extract often lack relevance to physicians conducting the examination. While some studies partially consider shape features, the features used don't align with the guidelines set forth in the TIRADS. When physicians

examine ultrasound images of thyroid nodules, they scrutinize both visual and textural nodule properties to classify it as benign or malignant. Geometric and morphological features embody the visual attributes of a nodule, providing the closest approximation to the features established by the gold standard in TIRADS. This has been observable across all the evaluated metrics, illustrating that the G-M features aptly mimic the visual characteristics defined in TIRADS. Key visual aspects of a thyroid nodule, such as margin irregularities and shape, directly correlate with the G-M features extracted in this research. The coupling of these features with machine learning resulted in the classification of thyroid nodules as either benign or malignant by considering the visual characteristics of the nodules (H4).

The evaluation of thyroid nodules in ultrasound images carries inherent subjectivity, leading to potential interobserver and intraobserver variability. Interobserver variability refers to the differences in assessments made by different observers (e.g., physicians), while intraobserver variability refers to the inconsistencies in evaluations made by the same observer at different times. Several factors contribute to these variabilities. First, the image quality and presentation of ultrasound can vary based on the settings used, patient anatomy, and nodule characteristics. Second, the interpretation of ultrasound images requires substantial expertise, and evaluations can differ based on the observer's training, experience, and individual biases. For example, one physician may perceive a nodule's echogenicity, shape, margin, or calcification differently from another, leading to discrepancies in assessments such as TIRADS classification. Existing scientific literature has documented these issues and is presented in the preceding chapters. Several studies [69-71] found substantial variability in TIRADS classification between observers, leading to differences in risk stratification of patients. Similarly, a study by Seifert et al. found that even when following different guidelines, there was

considerable interobserver variability in the interpretation of ultrasound features, which could impact management decisions [72].

The contributions made through this work focuses on the development of automated systems for the detection, segmentation, and classification of thyroid nodules in ultrasound images. Leveraging techniques such as geometric and morphological feature extraction, texture analysis, and machine learning, to mitigate subjectivity and variability in thyroid nodule analysis is what is needed to further augment and support current diagnostic practices. Moreover, these approaches also pave the way for more consistent and reliable diagnostics while also aiding in the tailoring of personalized therapeutic strategies. Deep learning techniques for the detection and segmentation of thyroid nodules also offers a solution to reduce variability. As these models learn from a large amount of data, they have the potential to outperform traditional methods in terms of consistency and reliability. Furthermore, all these techniques can augment the capabilities of physicians, especially in complex cases or in resource-limited settings where access to experienced clinicians may be limited.

4.1 Limitations and Future Scope

The general limitations of this work include the fact that the data used was limited. Further testing of these algorithms needs to be carried out in terms of data from various machines and manufacturers, imaging specialists, image resolution and dimensions, and geographic regions to ensure generalizability. Generalizability refers to the ability of an algorithm to work with different types of data inputs. Future scope of the work to include echogenic features as well as features describing calcified masses in the nodules needs to be considered. In doing so the technology has the potential to provide its user with end-to-end support through the further development of a single integrated system for thyroid nodule classification process. This would in turn provide them with

quantifiable data that they can understand and correlate to a reporting system such as TIRADS. Additionally, the developed systems need to be validated heavily in a clinical setting. This is aimed to be carried out in the near future by designing a study that measures observer variability with and without the use of the designed technologies. While subjectivity, interobserver, and intraobserver variability pose challenges in the detection and classification of thyroid nodules, advancements in image processing and machine learning offer promising avenues to address these issues. The contributions help in paving the way towards more reliable, consistent, and objective assessments of thyroid nodules in ultrasound images.

However, while these automated systems show great promise, it is essential to consider that they are tools designed to aid clinicians and are not meant to replace their expert judgement. Clinicians bring to bear not only their image interpretation skills but also their understanding of the patient's overall clinical context, which is currently beyond the scope of such automated systems.

Summary

Thyroid nodules are common and estimated to affect up to 65% of the general population with females being more prone to suffer from the disease than males. The vast majority are benign, with only 15% being malignant. Yet, differentiating between benign and malignant nodules is a critical task to avoid unnecessary invasive procedures on one side and to promptly treat malignancies on the other side. US imaging has emerged as the primary modality for thyroid nodule evaluation due to its non-invasiveness, cost-effectiveness, and accessibility.

US imaging allows clinicians to evaluate a nodule's size, structure, composition, and vascularity, all crucial for risk stratification and classification. Features such as irregular shape, microcalcifications, and taller-than-wide morphology raise suspicion for malignancy. However, the diagnostic accuracy of US depends heavily on the interpreting physician's experience and skill, leading to inherent issues with subjectivity and both interobserver and intraobserver variability.

The subjective nature of US interpretation means that different observers may evaluate the same nodule differently, creating interobserver variability. Additionally, intraobserver variability exists, whereby the same observer may evaluate the same nodule differently at different times. These variabilities contribute to diagnostic inconsistencies and may impact patient care decisions. Studies have shown that interobserver agreement on key ultrasound features like nodule echogenicity, margin irregularity, and the presence of microcalcifications can range from poor to moderate, emphasizing the potential for misdiagnosis or inconsistent risk stratification.

The work presented herein proposes a systematic approach for the detection, quantification and classification of thyroid nodules using US images aimed at reducing overall subjectivity and inter- and intraobserver variability in the diagnostic process. Techniques such as texture analysis, feature extraction, machine learning and deep learning were employed in the development of these approaches. The first step involved the use of autoregressive features together with Machine Learning to distinguish between the textures of the thyroid gland and surrounding structures in an attempt to focus the region to be examined. This was followed by the comparison of four deep learning algorithms to determine an appropriate one for detection and localization of thyroid nodules. The next step involved the use of autoregressive features and machine learning for the estimation and quantification of solid and cystic regions within US thyroid nodules. This additionally was aimed at aiding the physician in an objective method for dose calculation during radioiodine therapy. Finally, geometric and morphological features were extracted and coupled with machine learning techniques to classify the thyroid nodules. This final step especially takes into consideration the visual characteristics analyzed by physicians in accordance with TIRADS. Thus providing them with quantifiable evidence that supports the classification process.

The integration of computer diagnostic systems such as that proposed, is aimed at supporting physicians' decisions while enabling consistency. Future research will focus on refining these models, ensuring their robustness across diverse clinical contexts, machine manufacturers and varied data sources to determine how best to incorporate them into existing clinical workflows.

Zusammenfassung

Schilddrüsenknoten sind weit verbreitet und betreffen schätzungsweise bis zu 65 % der Allgemeinbevölkerung, wobei Frauen häufiger als Männer von der Krankheit betroffen sind. Die überwiegende Mehrheit ist gutartig, nur 15% sind bösartig. Die Unterscheidung zwischen gutartigen und bösartigen Knoten ist jedoch von entscheidender Bedeutung, um unnötige invasive Eingriffe auf der einen Seite zu vermeiden und bösartige Knoten auf der anderen Seite rechtzeitig zu behandeln. Die US-Bildgebung hat sich aufgrund ihrer Nichtinvasivität, Kosteneffizienz und Zugänglichkeit als primäre Modalität für die Beurteilung von Schilddrüsenknoten durchgesetzt.

Mit Hilfe der US-Bildgebung können Kliniker die Größe, Struktur, Zusammensetzung und Vaskularität eines Knotens beurteilen, die für die Risikostratifikation und -klassifizierung entscheidend sind. Merkmale wie eine unregelmäßige Form, Mikroverkalkungen und eine sogenannte Taller-than-wide-Morphologie, begründen einen Verdacht auf Malignität. Die diagnostische Genauigkeit des US hängt jedoch stark von der Erfahrung und den Fähigkeiten des interpretierenden Radiologen ab, was zu inhärenten Problemen mit der Subjektivität und der Variabilität zwischen und innerhalb der Beobachter führt.

Die subjektive Natur der US-Interpretation bedeutet, dass verschiedene Beobachter denselben Knoten unterschiedlich bewerten können, was zu einer Interobserver-Variabilität führt. Darüber hinaus gibt es eine Intraobserver-Variabilität, bei der ein und derselbe Beobachter denselben Knoten zu verschiedenen Zeitpunkten unterschiedlich bewertet. Diese Schwankungen tragen zu diagnostischen Unstimmigkeiten bei und können sich auf Entscheidungen der Patientenversorgung

auswirken. Untersuchungen haben gezeigt, dass die Übereinstimmung zwischen den Beobachtern bei wichtigen Ultraschallmerkmalen wie Echogenität des Knotens, Unregelmäßigkeit des Randes und Vorhandensein von Mikroverkalkungen schlecht bis mäßig sein kann, was das Potenzial für Fehldiagnosen oder eine inkonsistente Risikostratifikation bedingt.

In der hier vorgestellten Arbeit wird ein systematischer Ansatz für die Erkennung, Quantifizierung und Klassifizierung von Schilddrüsenknoten anhand von US-Bildern vorgestellt, der darauf abzielt, die allgemeine Subjektivität sowie die Variabilität zwischen und innerhalb der Beobachter im Diagnoseprozess zu verringern. Techniken wie Texturanalyse, Merkmalsextraktion, maschinelles Lernen und Deep Learning wurden bei der Entwicklung dieser Ansätze eingesetzt. In einem ersten Schritt wurden autoregressive Merkmale zusammen mit maschinellem Lernen verwendet, um zwischen den Texturen der Schilddrüse und der umgebenden Strukturen zu unterscheiden und so die zu untersuchende Region einzugrenzen. Anschließend wurden vier Deep-Learning-Algorithmen miteinander verglichen, um einen geeigneten Algorithmus für die Erkennung und Lokalisierung von Schilddrüsenknoten zu ermitteln. Im darauffolgenden Schritt wurden autoregressive Merkmale und maschinelles Lernen für die Einschätzung und Quantifizierung von soliden und zystischen Regionen innerhalb von US-Schilddrüsenknoten verwendet. Damit sollte dem Arzt zusätzlich eine objektive Methode zur Dosisberechnung bei der Radiojodtherapie zur Verfügung gestellt werden. Schließlich wurden geometrische und morphologische Merkmale extrahiert und mit Techniken des maschinellen Lernens gekoppelt, um die Schilddrüsenknoten zu klassifizieren. Dieser letzte Schritt berücksichtigt insbesondere die visuellen Merkmale, die von Ärzten gemäß TIRADS analysiert werden. Auf diese Weise können quantifizierbare Belege bereitgestellt werden, die den Klassifizierungsprozess unterstützen.

Die Integration computergestützter Diagnosesysteme, wie das hierin vorgestellte, zielt darauf ab, die Entscheidungen der Ärzte zu unterstützen und gleichzeitig Konsistenz zu gewährleisten. Künftige Forschungsarbeiten werden sich darauf konzentrieren, diese Modelle zu präzisieren und ihre Robustheit in verschiedenen klinischen Kontexten, bei unterschiedlichen Maschinenherstellern und unterschiedlichen Datenquellen zu gewährleisten, um eine optimale Integration in bestehende klinische Arbeitsabläufe zu eruieren.

Bibliography

[1] Benvenga, S., Tuccari, G., Ieni, A., & Vita, R. (2018). Thyroid gland: Anatomy and physiology. In Encyclopedia of Endocrine Diseases (pp. 382–390). Elsevier. https://doi.org/10.1016/b978-0-12-801238-3.96022-7

[2] Mitrou, P., Raptis, S. A., & Dimitriadis, G. (2011). Thyroid disease in older people. Maturitas, 70(1), 5–9. https://doi.org/10.1016/j.maturitas.2011.05.016

[3] Welker, M. J., & Orlov, D. (2003). Thyroid nodules. American Family Physician, 67(3), 559–566. https://www.ncbi.nlm.nih.gov/pubmed/12588078

[4] Yeung, M. J., & Serpell, J. W. (2008). Management of the solitary thyroid nodule. The Oncologist, 13(2), 105–112. https://doi.org/10.1634/theoncologist.2007-0212

[5] Zamora, E. A., Khare, S., & Cassaro, S. (2022). Thyroid Nodule. In StatPearls. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/pubmed/30571043

[6] Durante, C., Grani, G., Lamartina, L., Filetti, S., Mandel, S. J., & Cooper, D. S. (2018). The Diagnosis and Management of Thyroid Nodules: A Review. JAMA: The Journal of the American Medical Association, 319(9), 914–924. https://doi.org/10.1001/jama.2018.0898

[7] Grani, G., Sponziello, M., Pecce, V., Ramundo, V., & Durante, C. (2020). Contemporary Thyroid Nodule Evaluation and Management. The Journal of Clinical Endocrinology and Metabolism, 105(9), 2869–2883. https://doi.org/10.1210/clinem/dgaa322

[8] Meisinger, C., Ittermann, T., Wallaschofski, H., Heier, M., Below, H., Kramer, A., Döring, A., Nauck, M., & Völzke, H. (2012). Geographic variations in the frequency of thyroid disorders and thyroid peroxidase antibodies in persons without former thyroid disease within Germany. European Journal of Endocrinology / European Federation of Endocrine Societies, 167(3), 363–371. https://doi.org/10.1530/EJE-12-0111

[9] Haugen, B. R., Alexander, E. K., Bible, K. C., Doherty, G. M., Mandel, S. J., Nikiforov, Y. E., Pacini, F., Randolph, G. W., Sawka, A. M., Schlumberger, M., Schuff, K. G., Sherman, S. I., Sosa, J. A., Steward, D. L., Tuttle, R. M., & Wartofsky, L. (2016). 2015 American Thyroid Association Management Guidelines for Adult Patients with

Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid: Official Journal of the American Thyroid Association, 26(1), 1–133. https://doi.org/10.1089/thy.2015.0020

[10] Gharib, H., Papini, E., Garber, J. R., Duick, D. S., Harrell, R. M., Hegedüs, L., Paschke, R., Valcavi, R., Vitti, P., & AACE/ACE/AME Task Force on Thyroid Nodules. (2016). 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. Endocrine Practice: Official Journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists, 22(5), 622–639. https://doi.org/10.4158/EP161208.GL

[11] Grussendorf, M., Ruschenburg, I., & Brabant, G. (2022). Malignancy rates in thyroid nodules: a long-term cohort study of 17,592 patients. European Thyroid Journal, 11(4). https://doi.org/10.1530/ETJ-22-0027

[12] Pemayun, T. G. D. (2016). Current Diagnosis and Management of Thyroid Nodules. Acta Medica Indonesiana, 48(3), 247–257. https://www.ncbi.nlm.nih.gov/pubmed/27840362

[13] Popoveniuc, G., & Jonklaas, J. (2012). Thyroid nodules. The Medical Clinics of North America, 96(2), 329–349. https://doi.org/10.1016/j.mcna.2012.02.002

[14] Burman, K. D., & Wartofsky, L. (2016). Thyroid Nodules [Review of Thyroid Nodules]. The New England Journal of Medicine, 374(13), 1294–1295. https://doi.org/10.1056/NEJMc1600493

[15] Tamhane, S., & Gharib, H. (2016). Thyroid nodule update on diagnosis and management. Clinical Diabetes and Endocrinology, 2, 17. https://doi.org/10.1186/s40842-016-0035-7

[16] Elangovan, A., & Jeyaseelan, T. (2016). Medical imaging modalities: A survey. 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), 1–4. https://doi.org/10.1109/ICETETS.2016.7603066

[17] Kwak, J. Y., Han, K. H., Yoon, J. H., Moon, H. J., Son, E. J., Park, S. H., Jung, H. K., Choi, J. S., Kim, B. M., & Kim, E.-K. (2011). Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. Radiology, 260(3), 892–899. https://doi.org/10.1148/radiol.11110206

[18] Tessler, F. N., Middleton, W. D., Grant, E. G., Hoang, J. K., Berland, L. L., Teefey, S. A., Cronan, J. J., Beland, M. D., Desser, T. S., Frates, M. C., Hammers, L. W., Hamper, U. M., Langer, J. E., Reading, C. C., Scoutt, L. M., & Stavros, A. T. (2017). ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. Journal of the American College of Radiology: JACR, 14(5), 587–595. https://doi.org/10.1016/j.jacr.2017.01.046

[19] Russ, G., Bonnema, S. J., Erdogan, M. F., Durante, C., Ngu, R., & Leenhardt, L. (2017). European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS. European Thyroid Journal, 6(5), 225–237. https://doi.org/10.1159/000478927

[20] Priyanka, & Kumar, D. (2020). Feature Extraction and Selection of kidney Ultrasound Images Using GLCM and PCA. Procedia Computer Science, 167, 1722–1731. https://doi.org/10.1016/j.procs.2020.03.382

[21] Raghesh Krishnan, K., & Sudhakar, R. (2013). Automatic Classification of Liver Diseases from Ultrasound Images Using GLRLM Texture Features. Soft Computing Applications, 611–624. https://doi.org/10.1007/978-3-642-33941-7_54

[22] Iakovidis, D. K., Keramidas, E. G., & Maroulis, D. (2010). Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns. Artificial Intelligence in Medicine, 50(1), 33–41. https://doi.org/10.1016/j.artmed.2010.04.004

[23] Humeau-Heurtier, A. (2019). Texture Feature Extraction Methods: A Survey. IEEE Access, 7, 8975–9000. https://doi.org/10.1109/ACCESS.2018.2890743

[24] Shin, Y. G., Yoo, J., Kwon, H. J., Hong, J. H., Lee, H. S., Yoon, J. H., Kim, E.-K., Moon, H. J., Han, K., & Kwak, J. Y. (2016). Histogram and gray level co-occurrence matrix on gray-scale ultrasound images for diagnosing lymphocytic thyroiditis. Computers in Biology and Medicine, 75, 257–266. https://doi.org/10.1016/j.compbiomed.2016.06.014

[25] Micucci, M., & Iula, A. (2022). Recent Advances in Machine Learning Applied to Ultrasound Imaging. Electronics, 11(11), 1800. https://doi.org/10.3390/electronics11111800

[26] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and

Computer-Assisted Intervention – MICCAI 2015, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[27] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

[28] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z

[29] Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S. X., Ni, D., & Wang, T. (2019). Deep Learning in Medical Ultrasound Analysis: A Review. *Proceedings of the Estonian Academy of Sciences: Engineering*, *5*(2), 261–275. https://doi.org/10.1016/j.eng.2018.11.020

[30] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2980–2988. https://doi.org/10.1109/ICCV.2017.322

[31] Jan, J. (2019). Medical Image Processing, Reconstruction and Analysis: Concepts and Methods, Second Edition. CRC Press. https://play.google.com/store/books/details?id=vDL3DwAAQBAJ

[32] Weinmann, M., Jutzi, B., Mallet, C., & Weinmann, M. (2017). Geometric features and their relevance for 3d point cloud classification. ISPRS Annals of Photogrammetry Remote Sensing and Spatial Information Sciences, IV-1/W1, 157–164. https://doi.org/10.5194/isprs-annals-iv-1-w1-157-2017

[33] Zhao, J., Zheng, W., Zhang, L., & Tian, H. (2013). Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology. Health Information Science and Systems, 1, 5. https://doi.org/10.1186/2047-2501-1-5

[34] Kaur, J., & Jindal, A. (2012). Comparison of thyroid segmentation algorithms in ultrasound and scintigraphy images. International Journal of Computers & Applications, 50(23), 24–27. https://doi.org/10.5120/7959-0924

[35] China, D., Illanes, A., Poudel, P., Friebe, M., Mitra, P., & Sheet, D. (2019). Anatomical Structure Segmentation in Ultrasound Volumes Using Cross Frame Belief

Propagating Iterative Random Walks. IEEE Journal of Biomedical and Health Informatics, 23(3), 1110–1118. https://doi.org/10.1109/JBHI.2018.2864896

[36] Selvathi, D., & Sharnitha, V. S. (2011). Thyroid classification and segmentation in ultrasound images using machine learning algorithms. 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies, 836–841. https://doi.org/10.1109/ICSCCN.2011.6024666

[37] Mylona, E. A., Savelonas, M. A., & Maroulis, D. (2014). Automated adjustment of region-based active contour parameters using local image geometry. IEEE Transactions on Cybernetics, 44(12), 2757–2770. https://doi.org/10.1109/TCYB.2014.2315293

[38] Mylona, E. A., Savelonas, M. A., & Maroulis, D. (2014). Self-parameterized active contours based on regional edge structure for medical image segmentation. SpringerPlus, 3, 424. https://doi.org/10.1186/2193-1801-3-424

[39] Keramidas, E. G., Iakovidis, D. K., Maroulis, D., & Karkanis, S. (2007). Efficient and Effective Ultrasound Image Analysis Scheme for Thyroid Nodule Detection. Image Analysis and Recognition, 1052–1060. https://doi.org/10.1007/978-3-540-74260-9_93

[40] Maroulis, D. E., Savelonas, M. A., Iakovidis, D. K., Karkanis, S. A., & Dimitropoulos, N. (2007). Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images. IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society, 11(5), 537–543. https://doi.org/10.1109/titb.2006.890018

[41] Savelonas, M. A., Iakovidis, D. K., Legakis, I., & Maroulis, D. (2009). Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images. IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society, 13(4), 519–527. https://doi.org/10.1109/TITB.2008.2007192

[42] Garg, H., & Jindal, A. (2013). Segmentation of thyroid gland in ultrasound image using neural network. 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 1–5. https://doi.org/10.1109/ICCCNT.2013.6726797

[43] Narayan, N. S., Marziliano, P., Kanagalingam, J., & Hobbs, C. G. L. (2017). Speckle Patch Similarity for Echogenicity-Based Multiorgan Segmentation in Ultrasound

Images of the Thyroid Gland. IEEE Journal of Biomedical and Health Informatics, 21(1), 172–183. https://doi.org/10.1109/JBHI.2015.2492476

[44] Kollorz, E. K., Hahn, D. A., Linke, R., Goecke, T. W., Hornegger, J., & Kuwert, T. (2008). Quantification of thyroid volume using 3-D ultrasound imaging. IEEE Transactions on Medical Imaging, 27(4), 457–466. https://doi.org/10.1109/TMI.2007.907328

[45] Chang, C.-Y., Lei, Y.-F., Tseng, C.-H., & Shih, S.-R. (2008). Thyroid segmentation and volume estimation in ultrasound images. 2008 IEEE International Conference on Systems, Man and Cybernetics, 3442–3447. https://doi.org/10.1109/ICSMC.2008.4811830

[46] Iakovidis, D. K., Keramidas, E. G., & Maroulis, D. (2010). Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns. Artificial Intelligence in Medicine, 50(1), 33–41. https://doi.org/10.1016/j.artmed.2010.04.004

[47] Osman, A. (2013). Automated Evaluation of Three Dimensional Ultrasonic Datasets. Universitätsbibliothek der Universität Erlangen-Nürnberg. https://play.google.com/store/books/details?id=ZR93zwEACAAJ

[48] Poudel, P.; Gomes Ataide, E.J.; Illanes, A.; Friebe, M. (2018). Linear Discriminant Analysis and K-Means Clustering for Classification of Thyroid Texture in Ultrasound Images. '' in Proc. 40th Int. Conf. IEEE Eng. Med. Biol. Soc., Jul. 2018.

[49] Poudel, P., Illanes, A., Sheet, D., & Friebe, M. (2018). Evaluation of Commonly Used Algorithms for Thyroid Ultrasound Images Segmentation and Improvement Using Machine Learning Approaches. Journal of Healthcare Engineering, 2018, 8087624. https://doi.org/10.1155/2018/8087624

[50] Chen, J., You, H., & Li, K. (2020). A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. Computer Methods and Programs in Biomedicine, 185, 105329. https://doi.org/10.1016/j.cmpb.2020.105329

[51] Illanes, A., Esmaeili, N., Poudel, P., Balakrishnan, S., & Friebe, M. (2019). Parametrical modelling for texture characterization-A novel approach applied to ultrasound thyroid segmentation. PloS One, 14(1), e0211215. https://doi.org/10.1371/journal.pone.0211215

[52] Chang, C.-Y., Huang, H.-C., & Chen, S.-J. (2010). Automatic Thyroid Nodule Segmentation And Component Analysis In Ultrasound Images. Biomedical Engineering: Applications, Basis and Communications, 22(02), 81–89. https://doi.org/10.4015/S1016237210001803

[53] Haralick, R. M., Shanmugam, K., & Dinstein, I. 'hak. (1973). Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314

[54] Wu, C.-M., & Chen, Y.-C. (1992). Statistical feature matrix for texture analysis. CVGIP: Graphical Models and Image Processing, 54(5), 407–419. https://doi.org/10.1016/1049-9652(92)90025-S

[55] Galloway, M. M. (1975). Texture analysis using gray level run lengths. Computer Graphics and Image Processing, 4(2), 172–179. https://doi.org/10.1016/S0146-664X(75)80008-6

[56] Keramidas, E. G., Maroulis, D., & Iakovidis, D. K. (2012). TND: a thyroid nodule detection system for analysis of ultrasound images and videos. Journal of Medical Systems, 36(3), 1271–1281. https://doi.org/10.1007/s10916-010-9588-7

[57] Chapelle, O., Haffner, P., & Vapnik, V. N. (1999). Support vector machines for histogram-based image classification. IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council, 10(5), 1055–1064. https://doi.org/10.1109/72.788646

[58] Maroulis, D. E., Iakovidis, D. K., Karkanis, S. A., & Karras, D. A. (2003). CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames. Computer Methods and Programs in Biomedicine, 70(2), 151–166. https://doi.org/10.1016/s0169-2607(02)00007-x

[59] Ma, J., Wu, F., Jiang, T. 'an, Zhao, Q., & Kong, D. (2017). Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. International Journal of Computer Assisted Radiology and Surgery, 12(11), 1895–1910. https://doi.org/10.1007/s11548-017-1649-7

[60] Ying, X., Yu, Z., Yu, R., Li, X., Yu, M., Zhao, M., & Liu, K. (2018). Thyroid Nodule Segmentation in Ultrasound Images Based on Cascaded Convolutional Neural Network. Neural Information Processing, 373–384. https://doi.org/10.1007/978-3-030-04224-0_32

[61] Kumar, V., Webb, J., Gregory, A., Meixner, D. D., Knudsen, J. M., Callstrom, M., Fatemi, M., & Alizad, A. (2020). Automated Segmentation of Thyroid Nodule, Gland, and Cystic Components From Ultrasound Images Using Deep Learning. IEEE Access : Practical Innovations, Open Solutions, 8, 63482–63496. https://doi.org/10.1109/access.2020.2982390

[62] Nugroho, A., Nugroho, H. A., Setiawan, N. A., & Choridah, L. (2016). Internal content classification of ultrasound thyroid nodules based on textural features. Communications in Science and Technology, 1(2). https://doi.org/10.21924/cst.1.2.2016.25

[63] Zulfanahri, Nugroho, H. A., Nugroho, A., Frannita, E. L., & Ardiyanto, I. (2017). Classification of thyroid ultrasound images based on shape features analysis. 2017 10th Biomedical Engineering International Conference (BMEiCON), 1–5. https://doi.org/10.1109/BMEiCON.2017.8229106

[64] Pedraza, L., Vargas, C., Narváez, F., Durán, O., Muñoz, E., & Romero, E. (2015). An open access thyroid ultrasound image database. 10th International Symposium on Medical Information Processing and Analysis, 9287, 188–193. https://doi.org/10.1117/12.2073532

[65] Ataide, E. J. G., Schenke, S., Ghazzawi, S., Wüstemann, J., Illanes, A., Friebe, M., & Kreißl, M. C. (2020). Computer Aided Diagnosis: Initial Results for the Detection of Thyroid Nodules using US Images. NuklearMedizin 2020, 59, V5. https://doi.org/10.1055/s-0040-1708148

[66] Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing , 162, 94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013

[67] Nandamuri, S., China, D., Mitra, P., & Sheet, D. (2019). SUMNet: Fully Convolutional Model For Fast Segmentation of Anatomical Structures in Ultrasound Volumes. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 1729–1732. https://doi.org/10.1109/ISBI.2019.8759210

[68] Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018).

Attention U-Net: Learning Where to Look for the Pancreas. In arXiv [cs.CV]. arXiv. https://doi.org/10.48550/ARXIV.1804.03999

[69] Sahli, Z. T., Sharma, A. K., Canner, J. K., Karipineni, F., Ali, O., Kawamoto, S., Hang, J.-F., Mathur, A., Ali, S. Z., Zeiger, M. A., & Sheth, S. (2019). TIRADS Interobserver Variability Among Indeterminate Thyroid Nodules: A Single-Institution Study. Journal of Ultrasound in Medicine: Official Journal of the American Institute of Ultrasound in Medicine, 38(7), 1807–1813. https://doi.org/10.1002/jum.14870

[70] Phuttharak, W., Boonrod, A., Klungboonkrong, V., & Witsawapaisan, T. (2019). Interrater Reliability of Various Thyroid Imaging Reporting and Data System (TIRADS) Classifications for Differentiating Benign from Malignant Thyroid Nodules. Asian Pacific Journal of Cancer Prevention: APJCP, 20(4), 1283–1288. https://doi.org/10.31557/APJCP.2019.20.4.1283

[71] Liu, H., Ma, A.-L., Zhou, Y.-S., Yang, D.-H., Ruan, J.-L., Liu, X.-D., & Luo, B.-M. (2020). Variability in the interpretation of grey-scale ultrasound features in assessing thyroid nodules: A systematic review and meta-analysis. European Journal of Radiology, 129, 109050. https://doi.org/10.1016/j.ejrad.2020.109050

[72] Seifert, P., Görges, R., Zimny, M., & Schenke, S. (2019). Interobserver Agreement and Efficacy of Consensus Reading in Kwak-, EU-, ACR-TIRADS and ATA Guidelines for the Ultrasound Risk Stratification of Thyroid Nodules. NuklearMedizin 2019, 58, P28. https://doi.org/10.1055/s-0039-1683623

Danksagung

First and foremost, I would like to express my deepest gratitude to my esteemed supervisors, Prof. Dr. Michael Friebe and Prof. Dr. Michael Kreissl. Your endless patience, insightful criticisms, and stimulating ideas have been an anchor in my academic voyage. Your unwavering belief in my potential, even when I doubted myself, propelled me to surmount hurdles I previously thought insurmountable.

To my clinical advisors Dr. Simone Schenke, Dr. Sarvar Hahghi, Dr. Heiko Wissel, Dr. Sammy Ghazzawi and Dr. Jan Wüstermann who shared their time and expertise, your unique perspective and thorough understanding of practical aspects of our research provided an invaluable and essential bridge between theoretical principles and real-world application. You've helped me realize that there's much more to research than just what's found in books and articles.

I am also profoundly grateful to my technical advisors Dr. Alfredo Illanes and Dr. Axel Boese, who, with their expert guidance, steered my research in promising and fruitful directions. Your astute comments and constructive criticism have been indispensable in refining my work and thought process. Especially during my formative time when I was new to the field.

To my family, Beulah, Alaric, Pauline words fail to encapsulate the depth of my appreciation for your unyielding support and love. You have been a beacon of inspiration, your belief in me rock-solid through the highs and lows of this journey. The values of hard work, resilience, and passion you instilled in me have been the guiding stars of my academic pursuit. My brother Eldon whose unquestionable belief in me has been a beacon of hope when days were darkest.

My friends deserve a special mention for being my solace in times of stress and frustration. Lioba, Maria, Tome, Amit, Nikhil, Pranav, Laban, Joel, Carissa, Dylan, Esha, Srikanth, Nidhi, Gretchen, Meera, Sarah, Prabal, Nadine, Nikita, Nihat, Ali, Forough, Arwin, Marco, Nazila, Rutuja, Anna, Jenny and Muhannad. Your constant encouragement, understanding, and camaraderie have helped me maintain a sense of balance and sanity during this intense period of my life. Additionally, to everyone who contributed in various capacities to this endeavor, each of you has left a lasting imprint on my work and me. I am thankful for your collective wisdom, patience, and kindness. I am eternally indebted to all of you for making this journey not just possible, but truly rewarding and enlightening.

Lastly, to the person who believed in me the most. Although you are not with us your words and belief are the reason I am who I am today and have been able to achieve all that I have. Thank you for instilling in me the value of knowledge and always encouraging me to keep an open mind. Thank you for holding my hand, guiding me and teaching me kindness and empathy.

With heartfelt gratitude always,

Elmer Jeto Gomes Ataide

Eidesstattliche Erklärung

A Systematic Approach to the Detection, Quantification and Classification of Thyroid Nodules in Ultrasound Images using Image Computing, Machine and Deep Learning for Reduced Subjectivity and Inter- and Intraobserver Variability
im

der Klinik für Radiologie und Nuklearmedizin

der Medizinischen Fakultät
der Otto-von-Guericke-Universität Magdeburg
INKA- Application Driven Research

mit Unterstützung durch

Prof. Dr. rer. medic. Michael Friebe

Prof. Dr. med. Michael Kreissl

Dr. Alfredo Illanes

Dr. med. Simone Schenke

ohne sonstige Hilfe durchgeführt und bei der Abfassung der Dissertation keine anderen als die dort aufgeführten Hilfsmittel benutzt habe. Bei der Abfassung der Dissertation sind Rechte Dritter nicht verletzt worden.

Ich habe diese Dissertation bisher an keiner in- oder ausländischen Hochschule zur Promotion eingereicht. Ich übertrage der Medizinischen Fakultät das Recht, weitere Kopien meiner Dissertation herzustellen und zu vertreiben.

Magdeburg, den

Unterschrift

Description of educational background

Elmer Jeto Gomes Ataide was born on 08.12.1993. He completed his Bachelors in Medicine with a special focus on Health Information Management and Administration at the Manipal Academy of Higher Education in 2014. He was always passionate about developing medical technologies that could help people have access to quality diagnostic care irrespective of their location. With this in his sights he immediately pursued his Masters of Science in Health Informatics with a specialization in Healthcare IT Management at the Manipal Academy of Higher Education. During this time he worked on research areas such as improving the perception of Electronic Medical Record Systems among physicians and developing solutions for non-invasive blood glucose monitoring. After his Master's he worked as a full-time researcher in non-invasive blood glucose monitoring specifically aimed at improving diabetic foot health in rural areas of India. In the year 2017, he moved to Magdeburg-Germany to pursue his Doctoral studies under the supervision of Prof. Dr. Michael Friebe and Prof. Dr. Michael Kreissl at the Chair for Intelligent Katheters (INKA) at the Otto-von-Guericke University, Medical Faculty. Here he focused his research in the field of medical imaging and machine and deep learning. He worked towards developing solutions for ultrasound imaging using machine and deep learning that could provide physicians with objective decision support during the diagnostic process. He currently works at a Radiological AI company in Munich as the Head of Clinical Research developing medical AI products and clinical research strategies for clinical evaluation.

Elmer Jeto Gomes Ataide

List of All Publications

Gilberg, Leonard, Bianca Teodorescu, Leander Maerkisch, Andre Baumgart, Rishi Ramaesh, **Elmer Jeto Gomes Ataide**, and Ali Murat Koç. 2023. "Deep Learning Enhances Radiologists' Detection of Potential Spinal Malignancies in CT Scans." *NATO Advanced Science Institutes Series E: Applied Sciences* 13 (14): 8140.

Ataide, Elmer Jeto Gomes, Mathews S. Jabaraj, Alfredo Illanes, Simone Schenke, Axel Boese, Michael C. Kreissl, and Michael Friebe. 2022. "Thyroid Nodule Region Estimation Using Auto-Regressive Modelling and Machine Learning." *Current Directions in Biomedical Engineering* 8 (2): 588–91. https://doi.org/10.1515/cdbme-2022-1150

Ziegle, J., A. G. I. Manríquez, and **E. J. G. Ataide**. 2022. Methods and systems for thermal monitoring of tissue with an ultrasound imaging system. *US Patent 11,439,308*, issued 2022. https://patents.google.com/patent/US11439308B2/en. (US Patent)

Ataide, Elmer Jeto Gomes, Shubham Agrawal, Aishwarya Jauhari, Axel Boese, Alfredol Illanes, Simone Schenke, Michael C. Kreissl, and Michael Friebe. 2021. "Comparison of Deep Learning Algorithms for Semantic Segmentation of Ultrasound Thyroid Nodules." *Current Directions in Biomedical Engineering* 7 (2): 879–82. https://doi.org/10.1515/cdbme-2021-2224

Esmaeili, Nazila, Esam Sharaf, **Elmer Jeto Gomes Ataide**, Alfredo Illanes, Axel Boese, Nikolaos Davaris, Christoph Arens, Nassir Navab, and Michael Friebe. 2021. "Deep Convolution Neural Network for Laryngeal Cancer Classification on Contact Endoscopy-Narrow Band Imaging." *Sensors* 21 (23). https://doi.org/10.3390/s21238157.

Kuzhipathalil, Adarsh, Anto Thomas, Keerthana Chand, **Elmer Jeto Gomes Ataide**, Alexander Link, Annika Niemann, Sylvia Saalfeld, Michael Friebe, and Jens Ziegle. 2021. "A Machine Learning Approach Towards Fatty Liver Disease Detection in Liver Ultrasound Images." In *Bildverarbeitung Für Die Medizin 2021*, 86–91. Springer Fachmedien Wiesbaden.

Fritzsche, Holger, **Elmer Jeto Gomes Ataide**, Afshan Bi, Rohit Kalva, Sandeep Tripathi, Axel Boese, Michael Friebe, and Tim Gonschorek. 2020. "Innovative Hospital Management." *International Journal of Biomedical and Clinical Engineering* 9 (1): 33–47.

Fritzsche, Holger, **Elmer Jeto Gomes Ataide**, Axel Boese, and Michael Friebe. 2020. "Improved Patient Safety due to Catheter-Based Gas Bubble Removal during TURBT." *International Journal of Biomedical and Clinical Engineering* 9 (2): 1–11.

Gomes Ataide, Elmer Jeto, Holger Fritzsche, Marco Filax, Dinesh Chittamuri, Lakshmi Sampath Potluri, and Michael Friebe. 2020. "ENT Endoscopic Surgery and Mixed Reality." In *Biomedical and Clinical Engineering for Healthcare Advancement*, 17–29. IGI Global.

Gomes Ataide, Elmer Jeto, Nikhila Ponugoti, Alfredo Illanes, Simone Schenke, Michael Kreissl, and Michael Friebe. 2020. "Thyroid Nodule Classification for Physician Decision Support Using Machine Learning-Evaluated Geometric and Morphological Features." *Sensors* 20 (21). https://doi.org/10.3390/s20216110.

Köhler, Marcel, **Elmer Jeto Gomes Ataide**, Jens Ziegle, Axel Boese, and Michael Friebe. 2020. "Novel Assistive Device for Tomographic Ultrasound Neck Imaging vs. Freehand." *Current Directions in Biomedical Engineering* 6 (3): 28–31.

Ataide, Elmer Jeto Gomes, Jens Ziegle, Marco Kalmar, Sanchit Rathi, Shambhavi Shukla, Axel Boese, and Michael Friebe. 2019. "Feasibility and Initial Results of Assisted Ultrasound Scan Acquisition for Improved Tomographic Visualization." In *2019 IEEE 16th India Council International Conference (INDICON)*, 1–4.

Poudel, Prabal, Alfredo Illanes, **Elmer J. G. Ataide**, Nazila Esmaeili, Sathish Balakrishnan, and Michael Friebe. 2019. "Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction With Machine Learning Approaches." *IEEE Access* 7: 79354–65.

Publications included in the Dissertation

Publication 1

Poudel, Prabal, Alfredo Illanes, **Elmer J. G. Ataide**, Nazila Esmaeili, Sathish Balakrishnan, and Michael Friebe. 2019. "Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction With Machine Learning Approaches." *IEEE Access* 7: 79354–65. 10.1109/ACCESS.2019.2923547

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 29, 2019, accepted June 11, 2019, date of publication June 17, 2019, date of current version July 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923547

Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction With Machine Learning Approaches

PRABAL POUDEL, (Member, IEEE), ALFREDO ILLANES, ELMER J. G. ATAIDE, (Member, IEEE), NAZILA ESMAEILI, (Member, IEEE), SATHISH BALAKRISHNAN, AND MICHAEL FRIEBE, (Senior Member, IEEE)
Faculty of Medical Engineering, Otto-von-Guericke University, 39106 Magdeburg, Germany

Corresponding author: Prabal Poudel (prabal.poudel@ovgu.de)

This work supported by the Federal Ministry of Education and Research in the context of the INKA Project under Grant 03IPT7100X.

ABSTRACT The thyroid is one of the largest endocrine glands in the human body, which is involved in several body mechanisms like controlling protein synthesis, use of energy sources, and controlling the body's sensitivity to other hormones. Thyroid segmentation and volume reconstruction are hence essential to diagnose thyroid related diseases as most of these diseases involve a change in the shape and size of the thyroid over time. Classification of thyroid texture is the first step toward the segmentation of the thyroid. The classification of texture in thyroid Ultrasound (US) images is not an easy task as it suffers from low image contrast, presence of speckle noise, and non-homogeneous texture distribution inside the thyroid region. Hence, a robust algorithmic approach is required to accurately classify thyroid texture. In this paper, we propose three machine learning based approaches: Support Vector Machine; Artificial Neural Network; and Random Forest Classifier to classify thyroid texture. The computation of features for training these classifiers is based on a novel approach recently proposed by our team, where autoregressive modeling was applied on a signal version of the 2D thyroid US images to compute 30 spectral energy-based features for classifying the thyroid and non-thyroid textures. Our approach differs from the methods proposed in the literature as they use image-based features to characterize thyroid tissues. We obtained an accuracy of around 90% with all the three methods.

INDEX TERMS Medical imaging, support vector machine, artificial neural network, random forest classifier, texture classification, thyroid ultrasound.

I. INTRODUCTION

The thyroid is a butterfly shaped gland, one of the largest endocrine glands in the body, located below Adam's apple on the front of the neck. It is involved in several body mechanisms such as controlling protein synthesis, use of energy sources and controlling the body's sensitivity to other hormones. Due to these important functionalities, the thyroid is one of the important organs in the human body. However, it is susceptible to many diseases like Graves' (excessive production of thyroid hormones), subacute thyroiditis (inflammation of thyroid), thyroid cancer, goiter (thyroid swelling), etc [1]. In all of these cases, the size of the thyroid changes over time. So, it is essential to keep track of the thyroid size over time.

The associate editor coordinating the review of this manuscript and approving it for publication was Changsheng Li.

Ultrasound (US) imaging has been widely used for thyroid staging, as it is much safer and painless to use for the patients compared to other imaging modalities such as MRI which uses radio and magnetic waves, Computed Tomography (CT) which uses X-rays and Positron Emission Tomography (PET) which uses nuclear imaging technique [2]. Segmentation and volume computation of the thyroid have high clinical importance when it comes to the diagnosis and treatment of thyroid diseases. In this work, we will mainly focus on characterization of thyroid texture in an US image using three machine learning (ML) techniques. These approaches are Support Vector Machine (SVM), Artificial Neural Network (ANN) and Random Forest Classifier (RFC).

The features computed in this work for training the classifier are based on a novel texture characterization algorithm

published previously by our team [3]. A signal based parametrical approach using Autoregressive (AR) modelling has been proposed to characterize the thyroid texture using 30 AR spectral energy ratios based features that can distinguish between thyroid and non-thyroid regions. A simple clustering algorithm has been used to show the significance of the proposed AR-based features. In this new proposed work, we go further and use our robust textural features to train three different machine learning based approaches (SVM, ANN and RFC) that have already been used to segment US images in the literature. We show in this work that using the AR features together with the proposed classifiers the obtained results outperform other thyroid segmentation algorithm already presented in the literature.

The rest of the paper is organized as following: Section II presents the reviews on the related works on thyroid segmentation. Section III discusses about the novel feature extraction that we have used to extract signal based features from thyroid US images and the different texture classification methods. Section IV presents the results and compares our results with the ones from literature. Finally, Section V presents the discussion on the future works that we have planned as well as the conclusions that can be drawn from our work.

II. RELATED WORKS

Several approaches have been proposed on how to segment the thyroid in 2D US images. Zhao *et al.* [4] proposed several thyroid US segmentation approaches using edge detection, thresholding, region splitting and merging, watershed segmentation, active contour, graph theory, US image segmentation based on Ncut and segmentation based on improved normalized cut. Thyroid segmentation in 2D US and scintigraphy images using active contour without edges (ACWE), localized region based active contour and distance regularized level set were proposed by Kaur and Jindal [5]. China *et al.* [6] explored the possibilities of using the apriori information based on the US imaging physics and segmented the thyroid using Iterative Random Walks and Random Forest (IRWRF). Similarly, segmentation using a polynomial SVM [7], local region-based active contour [8], a boundary method and local binary patterns [9] for texture analysis and level-set active contours models [10] and [11] have been proposed. H. Garg and A. Jindal worked on feed-forward neural network (FNN) to segment the thyroid in US images [12]. Similarly, Echogenicity based Quantization (EBQ) and Joint Classification-Regression (JCR) which uses speckle related pixels and imaging artefacts as a source of information to perform multi-organ (i.e. thyroid, carotid artery, muscles and trachea) segmentation in thyroid US images were proposed by Narayan *et al.* [13].

Apart from segmentation in 2D images, several research works have been carried out to segment a fully 3D thyroid. A semi-automated approach to classify thyroid for volumetric quantification using geodesic active contour was proposed by Kollorz *et al.* [14]. Chang *et al.* [15] used a radial basis function (RBF) neural network to segment the blocks of thyroid gland. Similarly, a complete segmentation and analysis of 3D thyroid images was performed by Osman [16] by thresholding the voxel intensities and then connecting similar voxels to predict the thyroid regions. Poudel *et al.* [17] have used Active Contours without Edges (ACWE), Graph Cut (GC) and Pixel Based Classifier (PBC) to segment 2D thyroid images and later reconstructed them to compute a 3D thyroid.

Most of the above mentioned approaches involved thyroid segmentation using data-driven approaches which means that, the segmentation of thyroid was carrying out by directly operating over the pixel values in the US images. Similarly, several works have been proposed for thyroid nodule classification by characterizing the thyroid tissues. These works are based on computation of Statistical features [7], [18], [19], Spectral-based features [20], [21] and Higher Order statistics based features [22], [23]. The problem with using these data-driven approaches for feature computation is that, they are generally affected by the presence of speckle noise, low signal to noise ratio (SNR) and resolution in US images and even the pre-processing steps cannot get rid of these problems completely.

Similarly, most of the methods in the literature do not explore texture based features for thyroid segmentation. We believe this is due to the heterogeneous textural patterns within the thyroid US images [24] and thus a novel texture based feature extraction method should be devised to extract robust features which could be used to train the machine learning classifiers for thyroid segmentation.

As explained earlier, different machine learning based classifiers have been trained only using statistical, spectral and higher order statistical based features in the literature for thyroid texture classification. However, we have used a set of novel parametrical based features computed using AR modelling to classify the thyroid textures. To our knowledge, these features have not been used for training the machine learning based classifiers for thyroid texture classification. This is the main contribution of our work. We have used three widely used methods of texture classification from the literature and outperformed several other state-of-the art approaches which use different features compared to ours.

III. METHODS

This section is divided into four sections: database generation, features computation, texture classification and post-processing. In the first section, we will mainly discuss how the 2D US image datasets were acquired and how the texture patch database for training of the classifiers was prepared. The second section will present how the features were computed from the texture patches which were used for training of the classifiers and the third section presents the thyroid texture classification approach using SVM, ANN and RFC. Finally, the fourth section will explain a simple post-processing step that we have used to get rid of the over classified thyroid texture patches from the three trained classifiers.

A. THYROID DATASETS AND TEXTURE DATABASE GENERATION

A total of two 2D thyroid US datasets were used in this work. The first dataset (Dataset 1) consisted of six subjects with each subject containing between 53 and 189 2D thyroid US images. A total of 675 thyroid images with an image size of 760 x 500 pixels were used. This dataset was acquired by a medical expert in SurgicEye GmbH [35] and has been published and available in [25]. The second dataset (Dataset 2) consisted of sixteen subjects with each subject containing between 156 and 289 2D thyroid US images. The second dataset was obtained by a thyroid specialist medical doctor at University Clinic of Magdeburg, Germany and contains a total of 3, 370 thyroid US images with an image size of 760 × 1020 pixels. It has been presented in [26] and can be downloaded from http://opencas.webarchiv.kit.edu/?q=node/29 Along with the US images, we also acquired manually annotated ground truth images from the respective clinical experts who acquired the thyroid images. All the images were acquired using a General Electric (GE) Logiq E9 US machine equipped with Electromagnetic Tracking system. The acquired tracking data could be used for 3D reconstruction of segmented thyroid images and volume assessment over time.

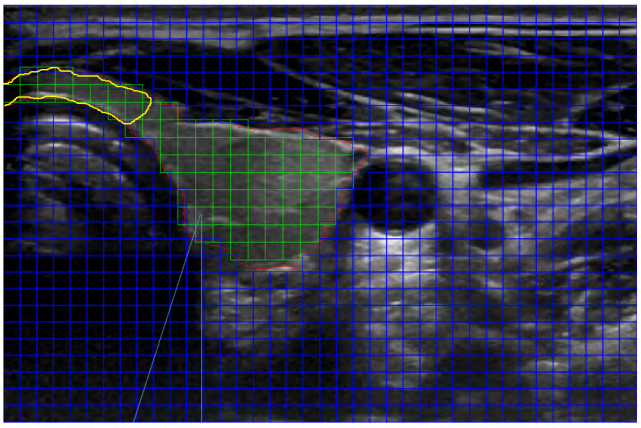


FIGURE 1. The figure represents the division of a 2D thyroid US images into smaller texture patches. In the figure, Green: Thyroid Patches, Blue: Non-Thyroid Patches, Red:Ground Truth and Yellow: Isthmus Region.

The two datasets were further processed to compute the features for training of the ML classifiers. Each image from the two datasets were first divided into non-overlapping texture patches of size 20 x 20 and following the ground truth, each patch was labelled either thyroid (=0) or non-thyroid (=1). The size of the texture patch was set in such a way that it captured important dynamical changes that allowed to involve a number of main frequency components that can help to spectrally differentiate thyroid and non-thyroid regions (see [3]). On top of that, it should also cover all the smaller regions inside the thyroid (for example the isthmus region as marked by yellow solid lines in Fig. 1).

For the labelling, a thresholding technique was used. Each pixel inside the patch was compared against the ground truth pixels. Hence, if a patch consisted of more than 70% (i.e. 280 pixels) of the total pixels, the patch was labelled as thyroid. Similarly, any patch that consisted of only black pixels (i.e. sum of all the pixel intensities inside a patch = 0) were not used as these patches could not be used to compute the features. A total of 90, 816 and 1, 791, 397 texture patches were computed from Dataset 1 and 2 respectively to prepare a final texture database. An example of separation of a 2D thyroid US image into texture patches is shown in Fig. 1. In the figure, the green patches represent the thyroid and the blue patches represent the non-thyroid patches. The thyroid patches are always present inside the thyroid region which is marked as red using the ground truth images.

B. FEATURES COMPUTATION

In this section, we will mainly discuss on how the features were computed from the thyroid images which were used for the training of the classifiers for thyroid texture classification. A detailed explanation on AR modelling, feature computation and prominent features selection have been explained in our recent work [3] but we will only introduce the main steps here. We used AR modelling to compute the features from the texture patches. The advantage of AR modelling is that the features are computed not directly from the image data (which in general contain speckle noise and have low SNR and contrast) like in Fast Fourier Transform based techniques, but using a parametrical version of the image data. This allows computing robust features in noisy images and less data compared to the standard data-driven methods.

First of all, the texture patches are converted into four different types of signals which capture the texture dynamics within the patch. The transformation from matrix to signal has been performed using ZigZag (obtained by following the rows direction) and Spiral transformation and also using their 90 degree rotated patch version (see Fig. 2).

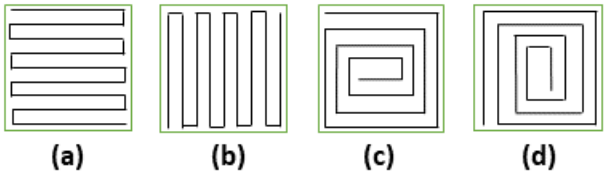


FIGURE 2. Conversion of texture patch to four different signals. ((a)ZigZag, (b)ZigZag 90 degree rotated, (c) Spiral and (d) Spiral 90 degree rotated respectively. Adopted from [3].

These signals were then decomposed into four narrowband signals (i.e. low, middle, high and total band frequency components - LF, MF, GF and TB respectively) by applying Continuous Wavelet Transformation (CWT). These signals represent the dynamic textural characteristics such as smoothness or roughness in the texture patches. A total of 16 narrowband signals were obtained which were modeled using a parametrical AR model [27]. A set of 30 different

FIGURE 3. Flowchart representing the entire feature extraction process.

features were computed from the AR parameters using the energy ratios between different frequency bands for each texture patch in the texture database. A detailed explanation on AR modelling, features computation and prominent features selection have been explained in [3]. A flowchart representing the entire feature computation process is shown in Fig. 3.

C. TEXTURE CLASSIFICATION

This section will present all the three algorithms that were used to classify the thyroid texture in US images.

1) SUPPORT VECTOR MACHINE (SVM)

In this work, SVM with radial basis function (rbf) (aka gaussian) kernel was used to classify the thyroid texture patches in US images. The features that were obtained from the feature extraction procedure were used to train the SVM. A total of 30 features were used to train the SVM classifier. The trained classifier was later used to test the input images by classifying the texture patches as thyroid or non-thyroid.

Let $x \in R^n$ be a vector of all the features extracted from the texture patches to be classified and let a scalar y denote its class label (i.e. whether the texture patch belongs to thyroid or not, $y \in \{0, 1\}$). Also, let $\{(x_i, y_i), i = 1, 2, 3, \ldots l\}$ be a set of l training data. For the simplest case, when the training patterns are linearly separable, there exist a linear function:

$$f(x) = W^T x + C \tag{1}$$

which separates the two different classes by a hyperplane:

$$f(x) = W^T x + C = 0 \tag{2}$$

where, C is the regularization parameter which controls the cost of misclassification on the training data.

However, there might exist many hyperplanes that maximize the separating margin between the two classes. The hyperplane that causes the largest separation between the different classes is computed by the SVM using minimizing the cost function [28]:

$$f(W) = \frac{1}{2} W^T W = \frac{1}{2} ||W^T||^2 \tag{3}$$

However, when the data are not linearly separable, a hyperplane cannot separate the data correctly. Thus, kernel functions are analysed to achieve this separation. In this work, a radial basis function (rbf) kernel is used which is given by:

$$k(x, z) = exp(-\frac{||x - z||^2}{2\sigma^2}) \tag{4}$$

where, $(\gamma = \frac{1}{2\sigma^2})$ is the kernel parameter that defines how far the influence of a single training example reaches. In other words, if the value of γ is low, then the far away points from the hyperplane carry more weights and if it is low, the nearer points carry more weights.

Using this kernel, all the features are sent as the input to the SVM classifier to train it. The features are represented in the vector form as:

$$x_i = [f_{i,1}, f_{i,2}, \ldots., f_{i,n}] \tag{5}$$

where $f_{i,n}$ is the n^{th} feature of the i^{th} texture patch.

These features are used as the training vector to train the SVM which is later used for testing the input images.The three parameters that could be optimized while using SVM are the kernel, C and the gamma γ. A grid search method with a 10-fold cross validation technique on the training data was employed to find the optimum parameters. We found that the SVM performed the best with 'rbf' kernel, $C = 0.7$ and $\gamma = 1.0$. A total of 75% training and 25% testing data were used to train and test the SVM classifier. The training and testing of SVM were carried out in Matlab 2017a using the Image Processing Toolbox.

2) ARTIFICIAL NEURAL NETWORK (ANN)

The classification of thyroid texture patches in US images was also done using ANN that is primarily an interconnected web of input nodes, hidden nodes and output nodes called artificial neurons.

The first step was to pre-process the data. The dataset was first split into dependent and independent variables. The independent variables consisted of the 30 features that were computed in the feature extraction section. The dependent variables consisted of data (represented as 1 or 0) that indicated whether the features belonged to thyroid or non-thyroid patches. Following this, the dataset was split into the training and testing sets by employing the train test split (75% and 25%) from scikit-learn model selection. Feature scaling was employed on the training and test sets to ensure that all the values were in the same range.

The ANN (Multi-layer Perceptron) was built with the Keras library using TensorFlow on the backend based on the Stochastic Gradient Descent Algorithm (SGD). SGD was used as an iterative method to adjust the weights and obtain a minimum cost function and hence an optimal neural network. The SGD is represented by the following equation:

$$Q(w) = \frac{1}{n} \sum_{i=1}^{n} Q_i(w)_i \tag{6}$$

where, $Q_i(w)$ is a loss function based on the training data indexed by i [29].

The optimization of the parameters of ANN involved some empirical analysis. After few experiments, we found that the ANN outperformed SVM and RFC. Hence, the following parameters were chosen as the optimum ones: 100 epochs, learning rate of 0.1, momentum of 0.9 and 4 layers. The 4 layers consisted of an input layer, two hidden layers (each of them containing 15 nodes) and an output layer. The Sequential module was used for the initialization of the network as a sequence of layers and the Dense module was used to build the layers.A Rectified Linear Unit (ReLU) activation function was used for the activation of the hidden layers. A sigmoid activation function was used for the output layers and is represented by:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

where, x = value of the weighted sums and e = Euler's number (= 2.71828) [30].

The classifier was saved after training the network with a batch size of 32 and 100 epochs. The trained ANN was used for testing the input images. The classification using ANN was carried out using Python 3.6 with the help of libraries such as Scikit-learn, Keras and TensorFlow.

3) RANDOM FOREST CLASSIFIER (RFC)

In our approach, we trained a random forest classifier for a binary classification problem, which classifies each of the patches extracted from the US images as thyroid and non-thyroid. RFC is basically a type of ensemble learning method which usually constructs a final classifier by using a set of M individual weak classifiers. In this case, these weak classifiers are the binary decision trees. A train-test split of 75% to 25% was used.

The input from the training data for each of the trees, $x \in \{1, \ldots.., M\}$ in the ensemble was created using bootstrapping of the samples (bagging) from the training dataset and randomly sampling the subset of the features supplied to each tree. Introducing this level of randomness helped this classifier in reducing to an extent, the dependency between training and testing data. Each tree is a collection of nodes N and features F, which aid to the final classification result. A decision tree is made up of a single parent node $N_{p,x}$ and multiple splitting nodes $N_{s,x,i} \forall i \in \{1, \ldots, k\}$ and leaf nodes $N_{l,x,j} \forall j \in 1, \ldots, p$. During the splitting of the nodes, the best split was not chosen based on all the features but a random subset of features from the training dataset.

All the leaf nodes inside a decision tree have a final probabilistic model $\phi_{x,j} \in [0, 1]$ associated with it. The final decision of a forest for each of the patches extracted from the US images were made by averaging the individual decisions ($\phi_{x,j}(p)$) from all the individual trees in the forest.

$$P^{RF}(y(p) = 1) = \frac{1}{M} \sum_{x=1}^{M} \phi_x(p) \tag{8}$$

We have used the most common and recognized method to train the classifiers [29], [31]. Just like ANN, the classification using RFC was carried out in Python 3.6 using Scikit-learn, Keras and TensorFlow libraries.

There are many parameters that can be optimized in RFC. However, we optimized only the 5 important parameters which were the depth of the trees, minimum number of samples required to split a node, minimum number of samples required at each leaf node, number of trees in the random forest and whether to use bootstrap or not. The optimum parameters that were obtained after using Randomized Search method were depth of 10, minimum samples at each leaf node of 2, minimum samples to split a node 4, 200 trees and using the bootstrap method for sampling the training data points.

D. POST-PROCESSING

The texture classification step produced some over-classified thyroid texture patches. Hence, to get rid of these over-classifications, a post-processing step was employed. A largest connected component analysis was performed on the classified texture patches. For that, the total number of texture patches were obtained by counting the patches that were classified as thyroid (i.e. the output label = 1). Then a threshold value was chosen empirically to identify the thyroid patches from the over-classified thyroid patches. The blocks of texture patches that contained more than the threshold amount of thyroid patches were considered to be thyroid and the rest were disregarded. Section IV C presents the results from before and after post-processing steps in details.

IV. RESULTS

A. EXPERIMENTAL SETUP

For the evaluation and quantitative and qualitative analysis of the proposed feature extraction and texture classification technique, we performed two-steps experiments. The two datasets were trained and tested separately. A total of 90, 816 and 1, 791, 397 texture patches corresponding to Dataset 1 and 2 respectively were used for this evaluation. Out of these patches, only 68, 112 patches were used for training and 22, 704 patches were used for testing in Dataset 1. Similarly, 1, 343, 548 patches were used for training and 447, 849 patches were used for testing in Dataset 2. In both datasets, to ensure there was no over-fitting while training of the classifiers, it was made sure that the training and testing processes did not involve images or texture patches from the same subjects. The training and testing processes involved the 75% and 25% of all the texture patches respectively.

The feature extraction part was performed using MATLAB 2017a and the training and testing of the classifiers was performed in Python 3.6. All the experiments were carried out using a Lenovo T430 ThinkPad Notebook with Intel Core i5-3320M CPU, NVIDIA NVS 5400 graphics card, 2.60 GHz processor and 8.00 GB RAM.

B. QUANTITATIVE ANALYSIS

For the quantitative analysis, we have compared our results with the approaches in state of art that used the same datasets. Similarly, we have also compared our approaches with other approaches but which do not use the same datasets. For the performance metric, we have used Dice's Coefficient (DC), Sensitivity (SE) and Specificity (SP).

DSC is a measure of how similar two objects are, which in our case is the computation of the overlap area between the ground truth images and classified thyroid texture patches. Similarly, **SE** is the measure of the proportion of actual positives that are correctly identified as such. **SP** is the measure of the proportion of actual negatives that are correctly identified as such. They can be computed using the following equations:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{9}$$

$$SE = \frac{TP}{TP + FN} \tag{10}$$

$$SP = \frac{TN}{TN + FP} \tag{11}$$

where, TP = True Positive (Thyroid Patches identified as Thyroid), FP = False Positive (Non-Thyroid Patches identified as Thyroid), TN = True Negative (Non-Thyroid Patches identified as Non-Thyroid) and FN = False Negative (Thyroid Patches identified as Non-Thyroid).

Using these performance metrics, we have presented the results of SVM, ANN and RFC and compared them with state of arts in the tables below. These comparisons are carried out in a 2-step procedure. The first step involved the comparison between all the approaches that use either Dataset 1 or 2 and in the second step, all the approaches were used for thyroid segmentation but using different datasets. The comparison of performance between SVM, ANN and RFC and state of arts are presented in Table 1 and 2 and Table 3 shows the comparison between different approaches that use different datasets. Table 4 summarizes all the parameters we used after the optimization process in SVM, RFC and ANN classifiers for texture classification.

TABLE 1. Performance comparison of SVM, ANN and RFC with state of art methods on Dataset 1.

Methods	DSC	SE	SP
ACWE [17]	0.805	-	-
GC [17]	0.745	-	-
PBC [17]	0.666	-	-
RFC - Volume Based [17]	0.855	-	-
CNN - Volume Based [17]	0.872	-	-
KMEANS [3]	0.897	0.950	0.700
SVM	**0.895**	**0.896**	**0.818**
ANN	**0.930**	**0.928**	**0.970**
RFC	**0.925**	**0.925**	**0.866**

Table 1 represents the comparison between the approaches we have used in our work with the works in [3] and [17] using Dataset 1. Active Contours without Edges (ACWE), Graph Cut (GC), Pixel based classifier (PBC), Random Forest Classifier (RFC) and Convolutional Neural Network (CNN) were used in [17] for thyroid segmentation. Out of these 5 approaches, the first three were non-machine learning (NML) based methods and the last two methods used machine learning (ML). However, these last two approaches were operated directly on 3D thyroid images. Similarly, kmeans (a simple clustering algorithm) was used in [3] to cluster and segment thyroid region in 2D thyroid US images.

TABLE 2. Performance comparison of SVM, ANN and RFC with state of art methods on Dataset 2.

Methods	DSC	SE	SP
IRWRF [6]	0.854	0.989	0.923
KMEANS [3]	0.869	0.890	0.620
SVM	**0.887**	**0.887**	**0.556**
ANN	**0.894**	**0.935**	**0.535**
RFC	**0.891**	**0.935**	**0.517**

TABLE 3. Performance analysis of different state of arts for thyroid segmentation using different Datasets.

Methods	DSC	SE	SP
EBQ [13]	0.839	0.955	0.889
JCR [13]	0.479	0.564	0.926
RBF [15]	0.512	0.874	0.560
FNN [12]	0.400	0.473	0.864

TABLE 4. Summary of all the optimized parameters used in SVM, RFC and ANN.

SVM	RFC	ANN
• kernel = 'rbf' • $C = 0.7$ • $\gamma = 1.0$	• tree depth = 10 • minimum samples to split leaf node = 2 • minimum samples to split a node = 4 • number of trees = 200 • bootstrap = 'true'	• epochs = 100 • learning rate = 0.1 • momentum = 0.9 • number of layers = 4

Similarly, Table 2 presents the comparison between our three approaches and Iterative Random Walks and Random Forest (IRWRF) from [6], a ML based and kmeans from [3], a NML based approaches using Dataset 2. We also present the results of thyroid segmentation using four other algorithms in Table 3. It presents the results using Echogenicity-based Quantization [13], Joint Classification-Regression (JCR) [13], RBF Neural Network (RBF) [15] and Feedforward Neural Network (FNN) [12] in terms of DSC, SE and SP using different thyroid US datasets. Despite the fact that these approaches use different datasets than we use, we present these results just to see how these algorithms perform in the domain of texture classification in thyroid US images.

All these metrics were computed using confusion matrix (CM) for each of the approaches used in our work. We present the CM for all the three algorithms when used on both the datasets below. In terms of TP, FN, FP and

TN, the CM can be represented as below. The CM were computed during the tests we carried out in the test sets which consisted of 22, 704 and 447, 849 texture patches in Dataset 1 and 2 respectively.

$$CM = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \tag{12}$$

Dataset 1:

$$\text{SVM} = \begin{bmatrix} 20317 & 2368 \\ 16 & 3 \end{bmatrix}$$

$$\text{ANN} = \begin{bmatrix} 20311 & 1576 \\ 22 & 795 \end{bmatrix}$$

$$\text{RFC} = \begin{bmatrix} 20290 & 1654 \\ 43 & 717 \end{bmatrix}$$

Dataset 2:

$$\text{SVM} = \begin{bmatrix} 397213 & 50600 \\ 34 & 2 \end{bmatrix}$$

$$\text{ANN} = \begin{bmatrix} 375914 & 26060 \\ 21333 & 24542 \end{bmatrix}$$

$$\text{RFC} = \begin{bmatrix} 374365 & 26133 \\ 22882 & 24469 \end{bmatrix}$$

From Table 1 and 2, we can see that all the three classifiers can classify the thyroid texture patches with better if not comparable accuracies. SVM has the lowest accuracy out of the three classifiers with a DC of 0.895 and 0.887 in Datasets 1 and 2 respectively. Similarly, ANN has the highest accuracy out of the three classifiers with a DC of 0.930 and 0.894 in Datasets 1 and 2 respectively. RFC produces almost the same accuracy as ANN with a DC of 0.925 and 0.891 in Datasets 1 and 2 respectively. These results can be visually accessed in the section below (see Section IV C). Similarly, all the three approaches outperformed ACWE, GC, PBC, RFC - Volume Based, CNN - Volume Based and KMEANS on Dataset 1 and KMEANS and IRWRF on Dataset 2 (except for KMEANS outperforming SVM on dataset 1). Apart from other methods, the tests with RFC and CNN - volume based were tested on the 3D thyroid volumes corresponding to Dataset 1 instead of individual 2D images.

We have also presented the performance matrices in terms of DSC, SE and SP from four different approaches in the literature such as EBQ, JCR, RBF and FNN despite the fact that they were tested on different datasets compared to what we are using in this work. These results are displayed in Table 3. Compared to these approaches too, SVM, ANN and RFC achieve better DSC and similar SE and SP in both the datasets. These results prove the robustness of the feature extraction process for thyroid texture classification.

Apart from the accuracy of classification, the feature extraction and training and testing of the approaches are fully automatic compared to ACWE, GC and PBC which use some level of human interaction. ACWE requires the user to draw an initial contour, GC requires the user to scribble the thyroid and non-thyroid region as a initialization process and PBC requires the users to click inside and outside of the thyroid regions to extract features from these regions. Also, the initializations are very important in these approaches as a wrong initialization could result in a misclassification of the different regions.

The computation time for feature extraction in our work is higher compared to the state of art techniques. This is mainly because we compute the wavelet spectrum for all the scales (or frequencies) in the LF, MF, HF and TB bands. An optimization step can be carried out to compute the spectrum at a scale that best represents these bands. Similarly, during AR modelling, instead of computing the power spectral densities (PSD) at all the frequency components in the complex plane, a set of non-repetitive frequency components could be chosen. On top of that, we have computed all the features using MATLAB which makes the process a lot slower. The optimization processes and the computation of these features in C++ could increase the frequency computation speed by a factor of 100. However, it is worth to mention that these features need to be computed only once and can be stored in a .csv file for training the networks in future. The time taken for classifying a new thyroid US image is however faster compared to the state of art methods. This makes it applicable for clinical use as the doctors and radiologist can just take a set of individual US images and segment the thyroid regions using the trained classifiers.

C. VISUAL ANALYSIS

The training of the three classifiers were followed by testing of individual images which were not part of the training set. An example of texture classification (first row) and segmentation (second row) on a total of 8 (4 from each dataset) different thyroid US images using SVM on Dataset 1 and 2 are shown in Fig. 4 and 5 respectively. Similarly, the results using ANN on Dataset 1 and 2 on the same images as in SVM are shown in Fig. 6 and 7 respectively. Fig. 8 and 9 show the results using RFC on Datasets 1 and 2 respectively. The images in the first row in all the figures from 4-9 show the results of texture classification using the trained classifiers and the images in the second row present the segmented thyroid regions after the post-processing step. In the figures, the green squares represent the 20 x 20 pixel texture patches classified as thyroid and the solid red line represents the ground truth region manually annotated by the expert clinicians. For testing purposes, we took the thyroid images from different locations with respect to the thyroid volume and from different patients.

The images from the first row in all the figures (i.e. Fig. 4, 5, 6, 7, 8 and 9) show the texture classification results from the trained classifiers and the images from the second row show the post-processed segmented thyroid region (marked with solid green lines). As evident in the

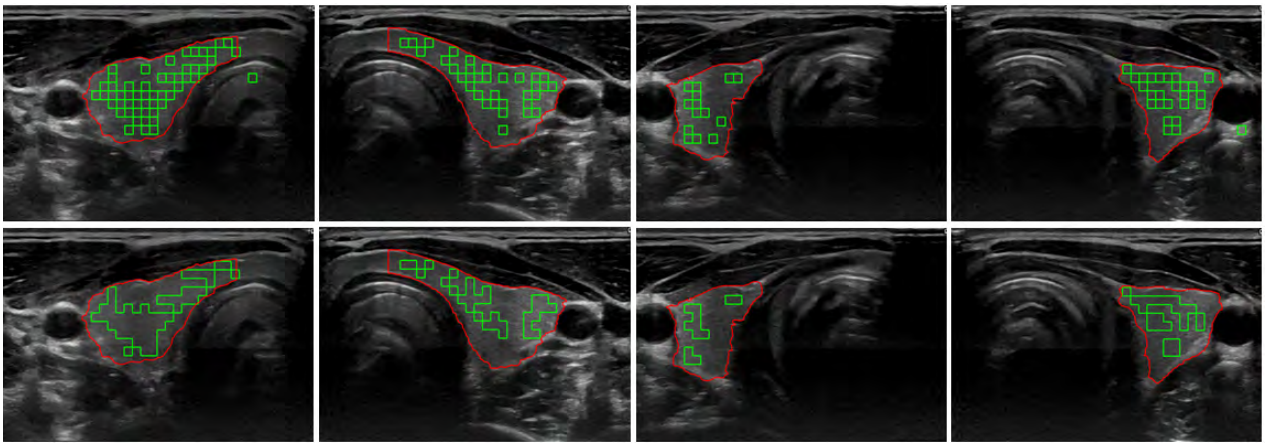


FIGURE 4. Examples of thyroid texture classification and segmentation using SVM and comparison with ground truth on Dataset 1.

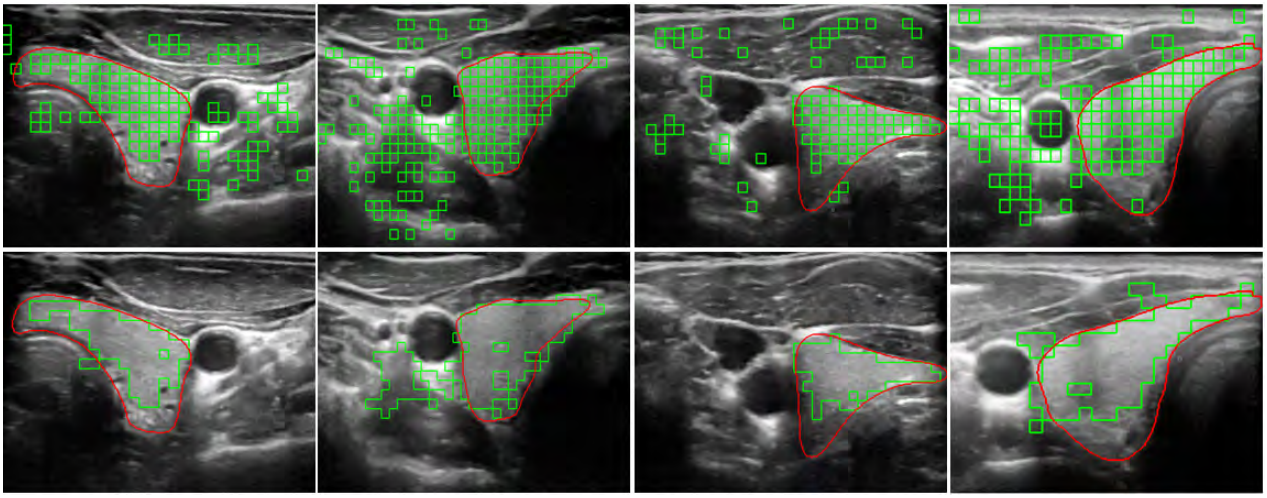


FIGURE 5. Examples of thyroid texture classification and segmentation using SVM and comparison with ground truth on Dataset 2.

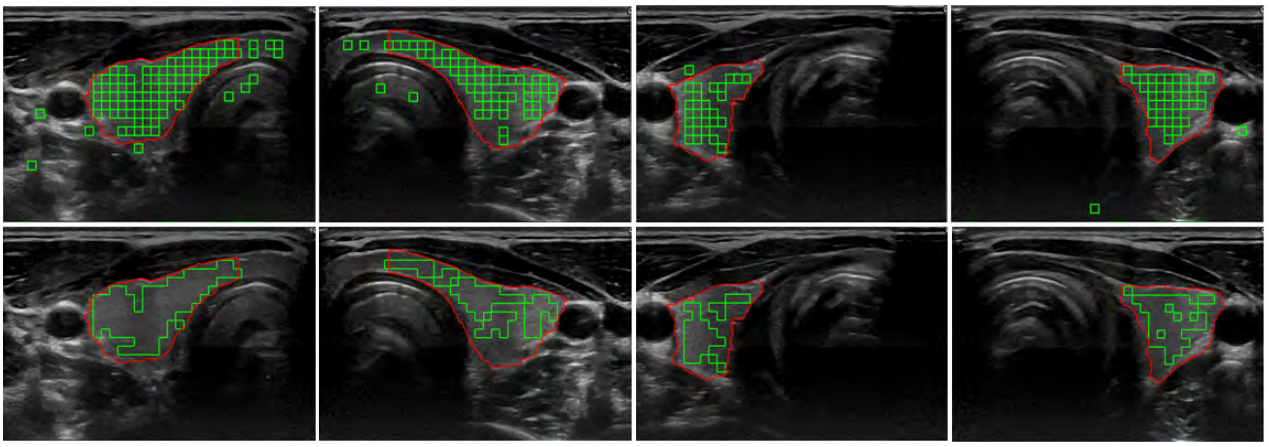


FIGURE 6. Examples of thyroid texture classification and segmentation using ANN and comparison with ground truth on Dataset 1.

figures, there are some over-classifications of texture patches as thyroid. Hence, a post-processing step was carried out to get rid of these over-classified texture patches.

The sample test images along with their ground truth have been shown above. From these test images, we can see that this way of texture classification obtains the larger thyroid

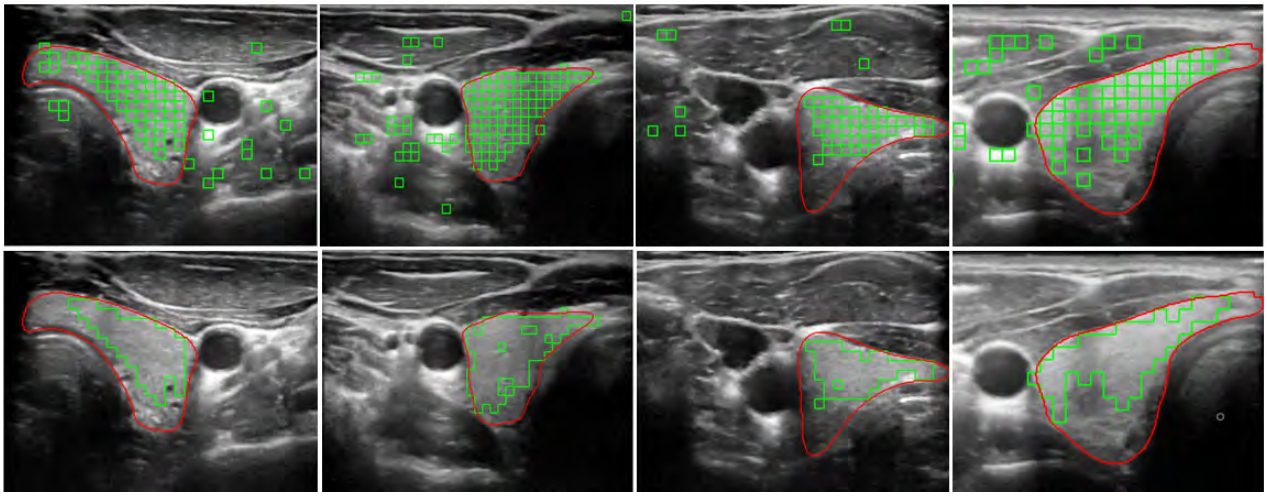


FIGURE 7. Examples of thyroid texture classification and segmentation using ANN and comparison with ground truth on Dataset 2.

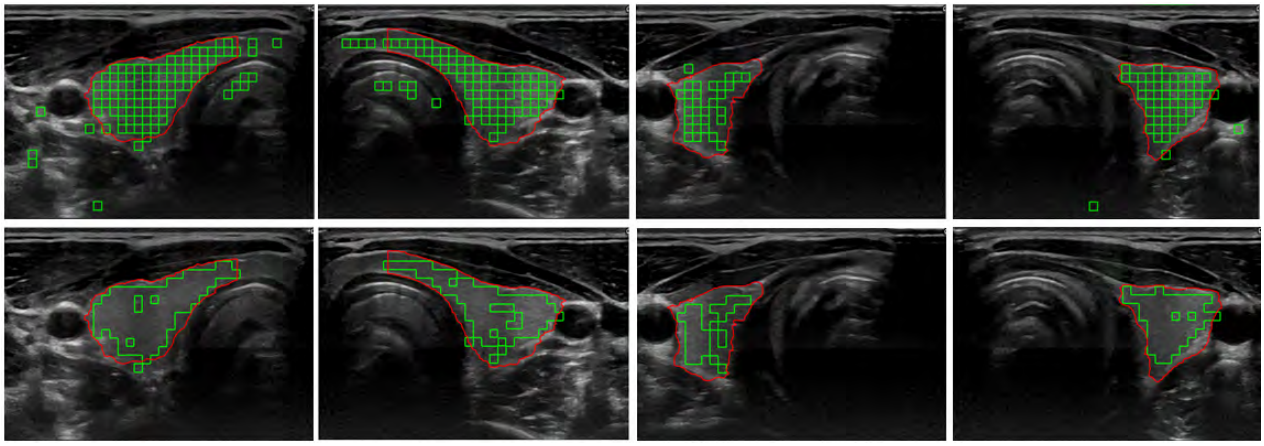


FIGURE 8. Examples of thyroid texture classification and segmentation using RFC and comparison with ground truth on Dataset 1.

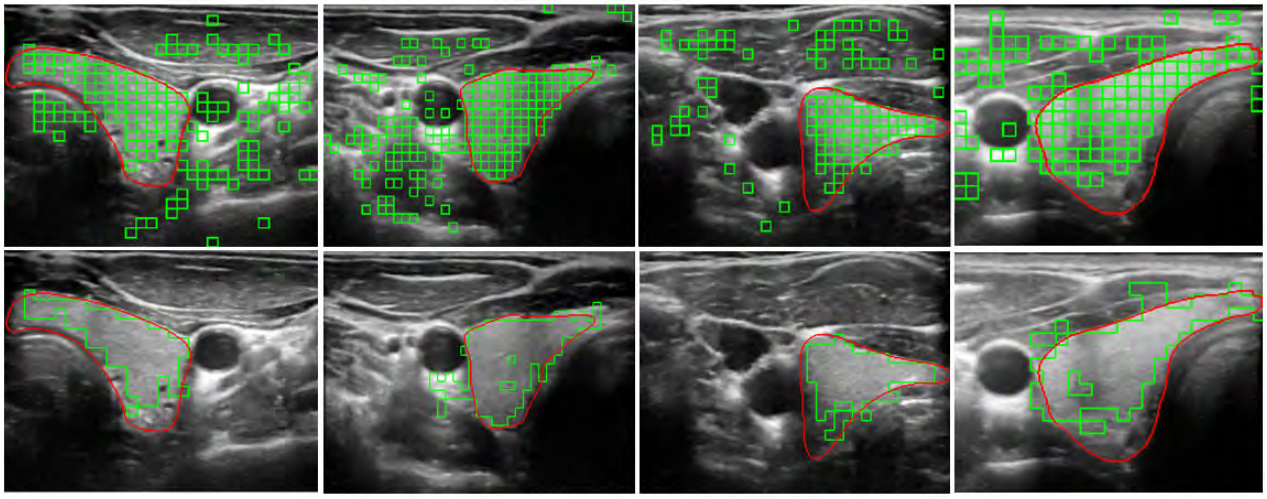


FIGURE 9. Examples of thyroid texture classification and segmentation using RFC and comparison with ground truth on Dataset 2.

region compared to the thyroid segmentation using ACWE, GC and PBC as they fail to segment the isthmus region inside the thyroid [17]. Despite classifying the regions in the isthmus, our approach achieves few under-classified results inside the thyroid. This problem could be solved by calculating more features (energy based, entropy based, statistical features, etc.) and using some extensive pre-processing techniques to choose the most prominent features like

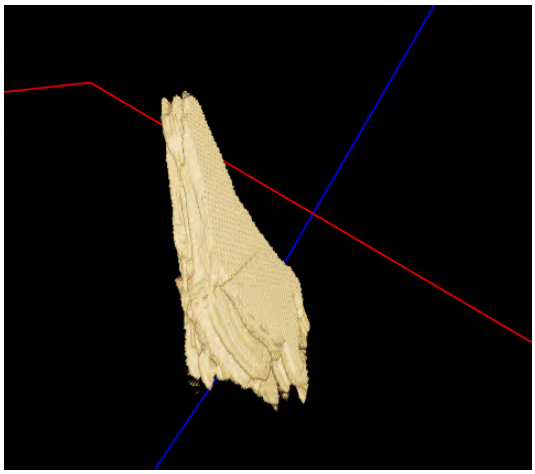


FIGURE 10. A 3D reconstructed thyroid after texture classification and segmentation using Imfusion.

Principal Component analysis [32] and Information Gain Techniques [33].

V. DISCUSSION AND CONCLUSION

In this paper, we have compared the three different machine learning techniques (SVM, ANN and RFC) for thyroid texture classification and segmentation. We computed the features for training of these classifiers using a very novel feature extraction technique. A signal based version of the US image was used and parametrically modelled to compute AR features. This transformation of the image to signal possesses many advantages compared to the image based analysis. With this way of treating the images, the extracted features are not affected by the presence of speckle noise, low contrast issues and low SNR in thyroid US images. This allows the classifiers to classify the thyroid region even in the smaller areas (for example in the isthmus of the thyroid) which was difficult using some of the state of art methods (ACWE, GC and PBC).

We also performed a comparison analysis between our approaches and various approaches in the literature. Two different comparison analysis were performed, first by comparing the performance on the same datasets and second on different datasets. In all of the tests, our approach outperformed the approaches in the literature in terms of DSC and had similar SE and SP. The results that we have obtained show a close correlation to the ground truth data. While the accuracy of training of the classifiers are similar, ANN slightly outperformed SVM and RFC. Our approaches were fully automated, so the user did not have to invest time in tracking the progress of the segmentation like in ACWE where the user had to stop the process and run it again, if the initialization of the contour was outside of the thyroid region. Similarly, in graph cut the user had to remove the over-classified regions after the segmentation and in PBC, the user had to make more clicks inside and outside the thyroid regions to get a better estimate of the features in order to train their decision trees.

One of the main drawbacks of the proposed approach is that it has only been evaluated with thyroid images from healthy subjects. In the future, we will explore how nodules can change the spectral behaviour in the US image. Similarly, we have used the images from a high-end machine (i.e. Logiq E9) for our task and the images from low-end machine might not have the same segmentation accuracy as we have shown in this work. Additionally, the classification of the texture patches always produces a non-smooth boarder in the segmented images. However, the clinical relevance could be established by training the classifiers on pathological images and the problem of the rough boarder in the segmented images could be solved by taking overlapping patches or by using a multi-resolution patch size and using the best size that produces a smooth appearance in the boarder region.

As future works, more features can be computed by not modelling the US images using AR model but by other methods as well like Bispectral model [34] and these features can be pre-processed by other pre-processing techniques such as Principal Component Analysis and Information Gain to select the prominent features. We could also combine the features from different modelling techniques and use them for the classification task. Similarly, the feature computation time can be reduced by optimizing the wavelet computation and AR modelling steps.

As mentioned above, the classified thyroid images can be reconstructed to a 3D volume as we also acquired the tracking data during the image acquisition phase. An example of the 3D reconstructed thyroid using Imfusion [36] after texture classification and segmentation is shown in Fig. 10.The Imfusion software allows the user to input all the binary images obtained from the segmentation as a video file along with the tracking matrices associated with each image frames. The reconstruction is then carried by using a technique called volumetric compounding where an interpolation is carried out between the corresponding image frames to fill the empty spaces. The 3D volume information can be used clinically by the medical experts to monitor the state of thyroid over time. Since most of the thyroid diseases involve change in the shape and volume of thyroid over time, the 3D reconstruction and volume computation has a clinical relevance.

ACKNOWLEDGMENT

The authors would like to thank General Electrics, USA, for providing us with the LogiqE9 Ultrasound Equipment to generate the Thyroid Ultrasound Images, and also would like to thanks to our clinical partners at the University of Magdeburg (Prof. C. Arens) for helping us obtain the Thyroid Ultrasound Datasets.

REFERENCES

[1] *Understanding Thyroid Problems—The Basics*. [Online]. Available: https://webmd.com/women/picture-of-the-thyroid

[2] T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*. New York, NY, USA: Elsevier, 2004, ch. 1.

[3] A. Illanes, N. Esmaeili, P. Poudel, S. Balakrishnan, and M. Friebe, "Parametrical modelling for texture characterization—A novel approach applied

to ultrasound thyroid segmentation," *PLoS ONE*, vol. 14, no. 1, Jan. 2019, Art. no. e0211215.
[4] J. Zhao, W. Zheng, L. Zhang, and H. Tian, "Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology," *Health Inf. Sci. Syst.*, vol. 1, no. 1, p. 5, Dec. 2013.
[5] J. Kaur and A. Jindal, "Comparison of thyroid segmentation algorithms in ultrasound and scintigraphy images," *Int. J. Comput. Appl.*, vol. 50, no. 23, pp. 1–4, Jan. 2012.
[6] D. China, A. Illanes, P. Poudel, M. Friebe, P. Mitra, and D. Sheet, "Anatomical structure segmentation in ultrasound volumes using cross frame belief propagating iterative random walks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 1110–1118, May 2019.
[7] D. Selvathi and V. S. Sharnitha, "Thyroid classification and segmentation in ultrasound images using machine learning algorithms," in *Proc. Int. Conf. Signal Process., Commun., Comput. Netw. Technol.*, Jul. 2011, pp. 836–841.
[8] N. H. Mahmood and A. H. Rusli, "Segmentation and area measurement for thyroid ultrasound image," *Int. J. Sci. Eng. Res.*, vol. 2, no. 12, pp. 1–8, Dec. 2011.
[9] E. G. Keramidas, D. K. Iakovidis, D. Maroulis, and S. Karkanis, "Efficient and effective ultrasound image analysis scheme for thyroid nodule detection," in *Proc. Int. Conf. Image Anal. Recognit.* Berlin, Germany: Springer, Aug. 2007, pp. 1052–1060.
[10] D. E. Maroulis, M. A. Savelonas, D. K. Iakovidis, S. A. Karkanis, and N. Dimitropoulos, "Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 5, pp. 537–543, Sep. 2007.
[11] M. A. Savelonas, D. K. Iakovidis, I. Legakis, and D. Maroulis, "Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 519–527, Jul. 2009.
[12] H. Garg and A. Jindal, "Segmentation of thyroid gland in ultrasound image using neural network," in *Proc. 4th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2013, pp. 1–5.
[13] N. S. Narayan, P. Marziliano, J. Kanagalingam, and C. G. Hobbs, "Speckle patch similarity for echogenicity-based multiorgan segmentation in ultrasound images of the thyroid gland," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 172–183, Jan. 2017.
[14] E. N. Kollorz, D. A. Hahn, R. Linke, T. W. Goecke, J. Hornegger, and T. Kuwert, "Quantification of thyroid volume using 3-D ultrasound imaging," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 457–466, Apr. 2008.
[15] C.-Y. Chang, Y.-F. Lei, C.-H. Tseng, and S.-R. Shih, "Thyroid segmentation and volume estimation in ultrasound images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1348–1357, Jun. 2010.
[16] A. Osman, "Automated evaluation of three dimensional ultrasonic datasets," Ph.D. dissertation, INSA de Lyon, Villeurbanne, France, 2013.
[17] P. Poudel, A. Illanes, D. Sheet, and M. Friebe, *J. Healthcare Eng.*, 2018.
[18] D. K. Iakovidis, E. G. Keramidas, and D. Maroulis, "Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns," *Artif. Intell. Med.*, vol. 50, no. 1, pp. 31–41, Sep. 2010.
[19] R. Koprowski, A. Korzyńska, Z. Wróbel, W. Zieleźnik, A. Witkowska, J. Małyszek, and W. Wójcik, "Influence of the measurement method of features in ultrasound images of the thyroid in the diagnosis of Hashimoto's disease," *Biomed. Eng. Online*, vol. 11, no. 1, p. 91, Dec. 2012.
[20] U. R. Acharya, S. V. Sree, M. M. R. Krishnan, F. Molinari, R. Garberoglio, and J. S. Suri, "Non-invasive automated 3D thyroid lesion classification in ultrasound: A class of ThyroScan systems," *Ultrasonics*, vol. 52, no. 4, pp. 508–520, Apr. 2012.
[21] U. R. Acharya, P. Chowriappa, H. Fujita, S. Bhat, S. Dua, J. E. W. Koh, L. W. J. Eugene, P. Kongmebhol, and K. H. Ng, "Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images," *Knowl.-Based Syst.*, vol. 107, pp. 235–245, Sep. 2016.
[22] U. R. Acharya, S. V. Sree, G. Swapna, S. Gupta, F. Molinari, A. Witkowska, and J. Suri, "Effect of complex wavelet transform filter on thyroid tumor classification in three-dimensional ultrasound," *Proc. Inst. Mech. Eng. H, J. Eng. Med.*, vol. 227, no. 3, pp. 284–292, Mar. 2013.
[23] U. Raghavendra, A. Gudigar, M. Maithri, A. Gertych, K. M. Meiburger, C. H. Yeong, C. Madla, P. Kongmebhol, F. Molinari, K. H. Ng, and U. R. Acharya, "Optimized multi-level elongated quinary patterns for the assessment of thyroid nodules in ultrasound images," *Comput. Biol. Med.*, vol. 95, pp. 55–62, Apr. 2018.
[24] P. Poudel, A. Illanes, and M. Friebe, "Ultrasound thyroid texture classification using a simple texture pattern characterization," in *Proc. Abstr. 51st Annu. Conf. German Soc. Biomed. Eng.*, vol. 62, Dresden, Germany, Sep. 2017.
[25] P. Poudel, E. J. G. Ataide, A. Illanes, and M. Friebe, "Linear discriminant analysis and K-means clustering for classification of thyroid texture in ultrasound images," in *Proc. 40th Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2018.
[26] T. Wunderling, B. Golla, P. Poudel, C. Arens, M. Friebe, and C. Hansen, "Comparison of thyroid segmentation techniques for 3D ultrasound," *Proc. SPIE*, vol. 10133, Feb. 2017, Art. no. 1013317.
[27] D. G. Manolakis, D. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*. New York, NY, USA: McGraw-Hill, 2000.
[28] S. Tsantis, D. Cavouras, I. Kalatzis, N. Piliouras, N. Dimitropoulos, and G. Nikiforidis, "Development of a support vector machine-based image analysis system for assessing the thyroid nodule malignancy risk on ultrasound," *Ultrasound Med. Biol.*, vol. 31, no. 11, pp. 1451–1459, Nov. 2005.
[29] A. Criminisi and J. Shotton, *Decision Forests for Computer Vision and Medical Image Analysis*. Berlin, Germany: Springer, 2013.
[30] L. Bottou, "Stochastic gradient descent tricks," in *Proc. Neural Netw., Tricks Trade*. Berlin, Germany: Springer, 2012, pp. 421–436.
[31] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[32] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.
[33] S. Lei, "A feature selection method based on information gain and genetic algorithm," in *Proc. Int. Conf. Comput. Sci. Electron. Eng.*. vol. 2, pp. 355–358, Mar. 2012.
[34] T. E. Hall and G. B. Giannakis, "Bispectral analysis and model validation of texture images," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 996–1009, Jul. 1995.
[35] *SurgicEye GmbH*. [Online]. Available: https://www.surgiceye.com/
[36] *ImFusion GmbH*. [Online]. Available: https://www.imfusion.de

PRABAL POUDEL (M'16) was born in Chitwan, Nepal, in 1993. He received the B.Sc. degree in electrical engineering and computer science from Jacobs University, Bremen, Germany, in 2014, and the M.Sc. degree in computer science from the University of Bonn, Germany, in 2016. He is currently pursuing the Ph.D. degree in medical engineering with Otto-von-Guericke University Magdeburg, Germany.

During the course of his B.Sc., he worked as a Research Assistant with Fraunhofer Mevis, Bremen, Germany, and recently he was a Visiting Researcher with General Electric Healthcare, Milwaukee, USA. His research interest includes medical image processing, computer vision, and machine learning.

ALFREDO ILLANES was born in Santiago, Chile, in 1978. He received the Electronic Engineering degree from UTFSM, Valparaiso, Chile, in 2002, and the M.Sc. degree in signal processing from the University of Nice Sophia Antipolis, France, in 2003, and the Ph.D. degree in signal processing (in the area of biosignal processing and modeling) from INRIA, Rennes, France, in 2008. From 2008 to 2015, he was an Assistant Professor with UACH, Valdivia, Chile, where his main research area was signal processing with applications in computer vision and vibratory processes. He is currently a Researcher with the Chair of Intelligent Catheter, Otto-von-Guericke University, Magdeburg, Germany. His main research interests include time-variant signal processing and modeling in vibratory and biological processes.

ELMER JETO GOMES ATAIDE (M'18) was born in 1993, in Goa, India. He received the B.Sc. degree in the medicine with a specialization in Health Information Administration and the M.Sc. degree in health informatics with a specialization in Healthcare IT Management from the Manipal Academy of Higher Education, Karnataka, India, in 2014 and 2016, respectively. He worked as a Research Assistant with the Diabetic Foot Clinic, Kastuba Medical College Hospital for one year and carried out his research on non-invasive glucose detection measuring core body temperature changes.

He is currently pursuing the Ph.D. degree in medical engineering with Otto-von-Guericke University, Magdeburg, Germany. His research interests include medical imaging, machine learning, and augmented reality in surgical suits and computer aided diagnosis.

NAZILA ESMAEILI (M'18) was born in Isfahan, Iran, in 1991. She received the B.Sc. degree in health information technology from the Isfahan University of Medical Sciences, Isfahan, in 2014 and the M.Sc. degree in medical systems engineering from Otto-von-Guericke University, Magdeburg, Germany, in 2018, where she is currently a Research and Development Engineer. From 2014 to 2016, she worked at the Informatics and IT Department of a private company, Isfahan. During the course of her M.Sc., she worked as a Research Assistant with INKA, Institute of Medical Technology, Magdeburg. Her research interests include medical image and signal processing, decision support systems and machine learning.

SATHISH BALAKRISHNAN was born in Coimbatore, India, in 1990. He received the B.E. degree in biomedical engineering from Anna University, India, in 2011, and the M.Sc. degree in biomedical computing from the Technical University of Munich (TUM), Germany, in 2017. He is currently a Research Assistant with INKA—Intelligente Katheter, Institute of Medical Engineering, Otto-von-Guericke University Magdeburg, Germany.

He has also worked as an Application Developer with IBM India Private Ltd., from 2011 to 2014. His research interests include medical image analysis, computer vision, machine learning, and deep learning, and developing robust and real-time solutions for ultrasound-based interventional procedures.

MICHAEL FRIEBE (M'05–SM'18) received the B.Sc. degree in electrical engineering, the M.Sc. degree in technology management from Golden Gate University, San Francisco, and the Ph.D. degree in medical physics, Germany. After the B.Sc. degree, he spent five years in San Francisco, as a Research and Design Engineer with an MRI and Ultrasound Device Manufacturer. He is a German citizen with expertise in diagnostic imaging and image guided therapies, as a Founder/Innovator/CEO/Invester, and a Scientist.

He is currently a Research Fellow with the Technical University of Munich, Munich, an Adjunct Professor with the Queensland University of Technology, Brisbane, and a Professor of image guided therapies with Otto-von-Guericke University, Magdeburg, Germany. He is a listed inventor of more than 80 patents and has authored more than 150 papers. He is a board member of four medical technology start-up companies, and investment partner of a MedTec investment-fund. From 2016 to 2018, he was a Distinguished Lecturer of the IEEE EMBC teaching innovation generation and MedTec entrepreneurship.

Publication 2

Ataide, Elmer Jeto Gomes, Shubham Agrawal, Aishwarya Jauhari, Axel Boese, Alfredol Illanes, Simone Schenke, Michael C. Kreissl, and Michael Friebe. 2021. "Comparison of Deep Learning Algorithms for Semantic Segmentation of Ultrasound Thyroid Nodules." *Current Directions in Biomedical Engineering* 7 (2): 879–82. https://doi.org/10.1515/cdbme-2021-2224

Elmer Jeto Gomes Ataide*, Shubham Agrawal, Aishwarya Jauhari, Axel Boese, Alfredo Illanes, Simone Schenke, Michael C. Kreissl and Michael Friebe

Comparison of Deep Learning Algorithms for Semantic Segmentation of Ultrasound Thyroid Nodules

Abstract: Ultrasound (US) imaging is used as a preliminary diagnostic tool for the detection, risk-stratification and classification of thyroid nodules. In order to perform the risk stratification of nodules in US images physicians first need to effectively detect the nodules. This process is affected due to the presence of inter-observer and intra-observer variability and subjectivity. Computer Aided Diagnostic tools prove to be a step in the right direction towards reducing the issue of subjectivity and observer variability. Several segmentation techniques have been proposed, from these Deep Learning techniques have yielded promising results. This work presents a comparison between four state of the art (SOTA) Deep Learning segmentation algorithms (UNet, SUMNet, ResUNet and Attention UNet). Each network was trained on the same dataset and the results are compared using performance metrics such as accuracy, dice coefficient and Intersection over Union (IoU) to determine the most effective in terms of thyroid nodule segmentation in US images. It was found that ResUNet performed the best with an accuracy, dice coefficient and IoU of 89.2%, 0.857, 0.767. The aim is to use the trained algorithm in the development of a Computer Aided Diagnostic system for the detection, risk-stratification and classification of thyroid nodules using US images to reduce subjectivity and observer variability.

Keywords: Ultrasound Imaging, Thyroid Nodules, Deep Learning, Image Processing, Computer Aided Diagnosis

https://doi.org/10.1515/cdbme-2021-2224

1 Introduction

Thyroid nodules are solid or cystic lumps in the thyroid gland, which can either be benign or malignant, and they are one of the most commonly diagnosed nodular lesions in the adult population [1]. Ultrasound (US) imaging is used as a preliminary diagnostic tool for the detection, risk-stratification and classification of thyroid nodules. It is used because of its availability, affordability and lack of ionizing radiation. In order to efficiently provide a risk-stratification and classification of a thyroid nodule, a physician first needs to detect the nodule. The current detection process is highly subjective in nature and there exists a high rate of inter-observer and intra-observer variability. This stems from the varied experience levels and visual perceptions of the physicians when it comes to detecting the nodules in US thyroid images.

Computer Aided Diagnosis is a solution that addresses the issue of inter-observer and intra-observer variability and subjectivity. The first step in this process is the detection of the thyroid nodule. This is also referred to as the segmentation of the nodule. Automatic segmentation of ultrasound images is a challenging task due to the heavy speckle noise, low contrast, and shadowing effects of ultrasound images.

Several segmentation techniques have been used like graph-based methods, deformable models, and classical methods. Kollorz et al. [1] presented an approach to semi-automatically segment 2D slices of collected 3D US volumes using power watersheds independent of the type of the nodule. Keramidas et al. [2] presented a Thyroid Boundaries Detection algorithm for detection of the thyroid area and later applied KNN to classify the small patches out from the sliding window based on the Local Binary Patterns texture feature. Even though it was an efficient technique, it could only output rough shapes and locations where the nodules were present. Lakovidis et al. [3] applied active contour methodologies for fine detection of thyroid nodules. Chan

***Corresponding author: (Elmer Jeto) (Gomes Ataide):** INKA-Healthtech Innovation Laboratory & Department of Radiology and Nuclear Medicine, Otto-von-Guericke University Magdeburg, Medical Faculty, Leipziger Str. 44, Magdeburg, Germany, e-mail: elmer.gomesataide@ovgu.de
2nd Author (Shubham) (Agrawal), 3rd Author (Aishwarya) (Jauhari): Faculty of Information Technology, Otto-von-Guericke University, Magdeburg, Germany
4th Author (Axel) (Boese), 5th Author (Alfredol) (Illanes): INKA-Healthtech Innovation Laboratory, Otto-von-Guericke University Magdeburg, Medical Faculty, Magdeburg, Germany
6th Author (Simone) (Schenke), 7th Author (Michael C.) (Kreissl): Department of Radiology and Nuclear Medicine, Division of Nuclear Medicine, Otto-von-Guericke University Magdeburg, Medical Faculty, Magdeburg, Germany
8th Author (Michael) (Friebe): INKA-Healthtech Innovation Laboratory, Otto-von-Guericke University Magdeburg, Medical Faculty, Magdeburg, Germany

and Vesse [4] suggested Active Contour without Edges technique based on the level-set method. Deep learning models have also proven to be efficient in segmentation problems and have achieved noteworthy results. Some deep learning-based segmentation methods have also been used for medical images and achieved exceptional improvements by using deep convolutional neural networks (DCNN)[5], fully convolutional neural networks (FCN)[6], and its variants[7].

This paper presents a comparison between four state-of-the-art (SOTA) algorithms developed for the purpose of image segmentation. Each algorithm architecture was adapted for thyroid US images. The results obtained are compared and the best in terms of its overall performance is highlighted. The purpose if doing this is to determine which of these methods can be used most effectively for the development of a complete Computer Aided Diagnostic system for the detection, risk-stratification and classification of thyroid nodules using US images.

2 Methods

This study used US image data of thyroid nodules collected at the Department of Radiology and Nuclear Medicine, Division of Nuclear Medicine, Otto-von-Guericke University Magdeburg, Medical Faculty. A total of 1011 images were used in this study. Each image consisted of a minimum of one thyroid nodule scanned in both transverse and longitudinal planes.

2.1 Data Pre-processing and Augmentation

The images were first pre-processed in order to remove unnecessary background information and noise and then resized and cropped. This was followed by dilation using a kernel of size 15x15. The contours were identified using CHAI_APPROX_SIMPLE algorithm on the processed output image from the previous step and reshaped the image based on the coordinates of the bounding rectangle of the contour. The non-square images were then transformed into square images of dimensions 256*256. The pre-processed image was then centred on its mean and normalized. These processed images were then used for training four SOTA segmentation algorithms. Data augmentation in terms of horizontal and vertical shifts, rotation at angles in the range of 5 to 15 degrees, different affine transformations, and grey value variations were employed. Additionally, random elastic deformations used in the training samples delivered better generalization results. The final augmented number of images was 6066. A train-test split of 75%-25% was used during the experiments.

2.2 Experimental Setup

Four SOTA sematic segmentation algorithms were selected to be tested and compared in this study. These were UNet [6], SUMNet [8], ResUNet [9] and Attention UNet [10]. The architectures for these networks were adapted to fit the data specifically. The details are given below.

2.2.1 UNet

The network architecture was designed as per the specifications given in [6]. In order to make the model more robust, Gaussian noise was applied in the input along with histogram equalization to make the network contrast independent. Stochastic Gradient Descent was used as an optimizer along with Binary cross entropy as a loss function. The network was trained for 350 epochs with a batch size of 16 and early stoppage.

2.2.2 SUMNet

The network architecture was designed as per the specifications given in [8]. Adam optimizer was used along with a weighted combination of Dice loss and binary cross-entropy as loss functions. Additionally, a learning rate of 1e-4 and a beta1=0.99 was set for the network. The network was trained for 350 epochs with a batch size of 16 and early stoppage.

2.2.3 ResUNet

The architecture of the network was designed according to the specifications given in [9]. Similarly to UNet, Gaussian noise was applied in the input, along with histogram equalization so as to make the network contrast independent and linearly scaled to obtain the Gaussian distribution. The network was trained using an Adam optimizer and a learning rate of 1e-4 and beta1= 0.99. Binary cross-entropy was used as a loss function. The network was trained for 350 epochs with a batch size of 32 and early stoppage.

2.2.4 Attention UNet

The architecture of the network was designed according to the specifications given in [10]. The model was trained using the Adam optimizer, along with the momentum=0.99, batch-normalization after each convolutional layer, and deep supervision. The model used binary cross-entropy and Dice loss as loss functions. Gating parameters were initialized so that attention gates can pass through every feature vector at all spatial locations. The network was trained for 350 epochs with a batch size of 32 and early stoppage.

2.3 Evaluation

The evaluation of the output obtained from each algorithm was done using the following three techniques. These are highlighted as follows:

Accuracy: Can be defined as the percentage of true positive outcomes against total outcomes.

Dice Coefficient: Dice coefficient can be defined as (2 * area Of overlap between predicted output and ground truth)/total number of pixels in both the images. Is positively correlated and its value also ranges from 0 to 1.

Intersection over Union (IoU): Intersection-Over-Union (IoU), also called the Jaccard Index, is the most commonly used metric to measure the performance of semantic segmentation. It is defined as the (area of overlap between the predicted output and the ground truth)/(area of union between the predicted output and the ground truth).

3 Results & Discussion

Table 1 shows the comparison of the results for each of the selected evaluation metrics for US thyroid nodule segmentation. Out of all the proposed architectures, ResUNet yields the highest value for all the metrics, followed by Attention UNet, SUMNet, and UNet. Figure 1 shows the training and validation loss curves for ResUNet. The training curve depicts a slow descent to a reduced loss value. This ensured proper training of the network while preventing over-fitting. The validation curve depicts a rapid descent to a low loss value and is maintained through the remaining epochs.

Table 1: Comparison of results obtained from all four algorithms for the selected evaluation metrics for the semantic segmentation of US thyroid nodules

Model/Metrics	Accuracy (%)	Dice Coefficient	IoU
UNet	79.7	0.766	0.635
SUMNet	83.1	0.795	0.672
ResUNet	89.2	0.857	0.767
Attention UNet	85.4	0.821	0.720

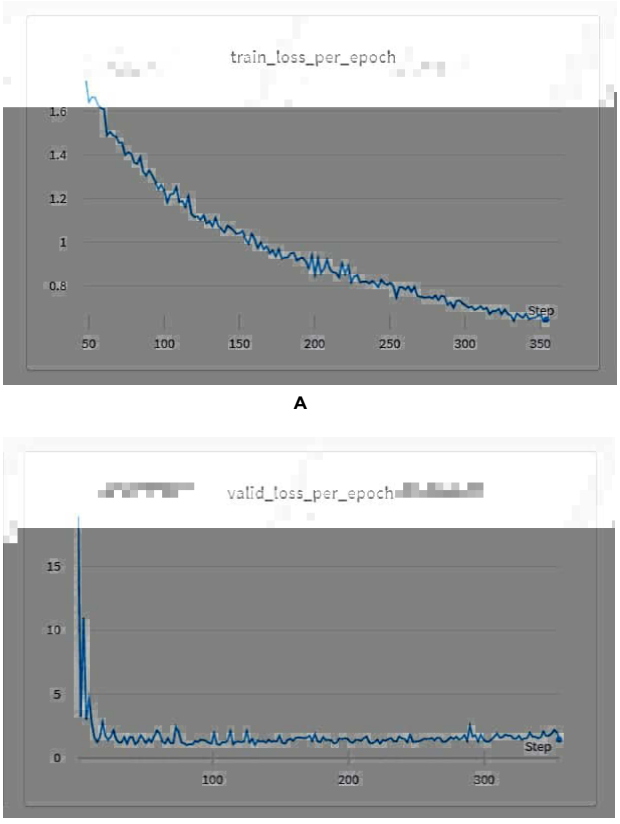


Figure 1: Training (A) and validation (B) loss curves for the ResUNet

Figure 2 shows the segmentation results from two different nodules as examples. The segmentation results are given for each of the algorithms used along with the original images and ground truths. It can be seen that the segmentation outputs reflect the results calculated by the metrics. Among the four selected algorithms, ResUNet results in a

segmentation that is closest to the ground truth compared to the remaining three algorithms.

4 Conclusion

The segmentation of thyroid nodules is a crucial task as it acts as the first step towards the overall classification. This work presents the evaluation results for the segmentation of thyroid US images using four different architectures. From the results, it can be observed that ResUNet architecture outperforms other architectures with a high value for all the metrics used for the evaluation of the model. One of the major reasons for superior results obtained using ResUNet is the presence of residual blocks of convolutional layers that enables consistent training of the network as the depth of the network increases. Residual blocks also enable easy flow of gradients to the initial layers using skip connections and this helps in training the network fast in the early stages. The next steps would be testing the developed ResUNet architecture with more data and its subsequent integration in a Computer Aided Diagnostic system for the detection, risk-stratification and classification of thyroid nodules that can be used by the physicians to reduce overall diagnostic subjectivity.

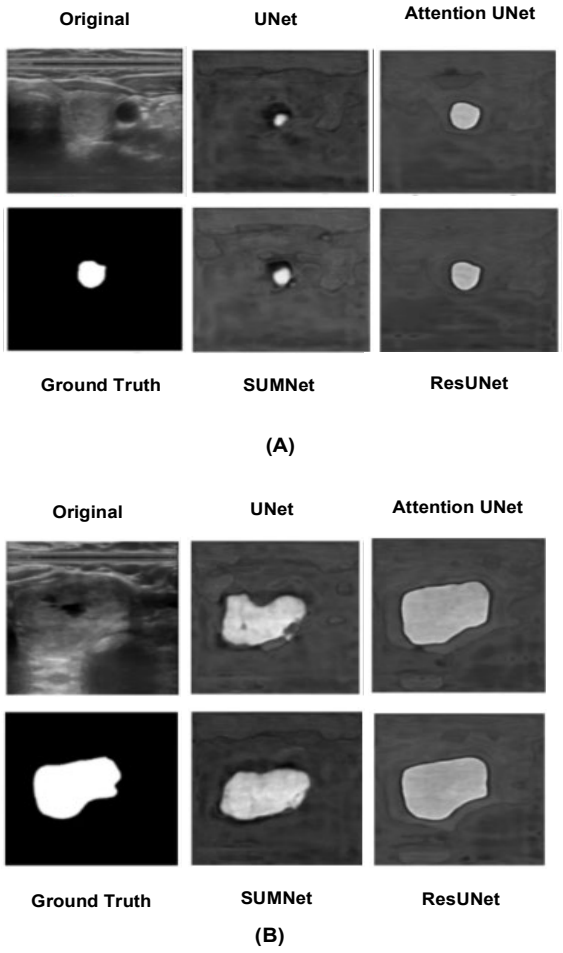


Figure 2: Segmentation outputs from each of the algorithms along with the original image and ground truths

Author Statement

The authors state no conflict of interest. Informed consent was taken. Ethical approval was obtained (RAD362-16/19) at the Otto-von-Guericke University Hospital, Magdeburg.

References

[1] E. Kollorz, E. Angelopoulou, M. Beck, D. Schmidt, and T. Kuwert, "Using Power Watersheds to Segment Benign Thyroid Nodules in Ultrasound Image Data," Bildverarbeitung für die Medizin 2011, pp. 124–128, 2011.

[2] Keramidas EG, Iakovidis DK, Maroulis D, Karkanis S (2007) Efficient and Effective Ultrasound Image Analysis Scheme for Thyroid Nodule Detection[C]. International Conference on Image Analysis and Recognition. Springe-Verlag, 4633: 1052-1060.

[3] Iakovidis DK, Savelonas MA, Karkanis SA, Maroulis DE(2006) Segmentation of Medical Images with Regional Inhomogeneities[C]. International Conference on Pattern Recognition. IEEE Computer Society, 3: 976-979.

[4] Chan TF, Vesse LA (2001) Active Contours Without Edges. IEEE Trans.Image Processing. 7: 266-277.

[5] J. Ma, F. Wu, T. Jiang, Q. Zhao, and D. Kong, "Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks," International Journal of Computer Assisted Radiology and Surgery, vol. 12, no. 11, pp. 1895–1910, 2017.

[6] O. Ronneberger, "Invited Talk: U-Net Convolutional Networks for Biomedical Image Segmentation," Informatik aktuell, pp. 3–3, 2017.

[7] Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," 2016 Fourth International Conference on 3D Vision (3DV), 2016.

[8] S. Nandamuri, D. China, P. Mitra, and D. Sheet, "SUMNet: Fully Convolutional Model For Fast Segmentation of Anatomical Structures in Ultrasound Volumes," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019.

[9] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 162, pp. 94–114, 2020.

[10] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich,Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert: Attention U-Net: Learning Where to Look for the Pancreas. 2018 Computer Vision and Pattern Recognition.

Publication 3

Ataide, Elmer Jeto Gomes, Mathews S. Jabaraj, Alfredo Illanes, Simone Schenke, Axel Boese, Michael C. Kreissl, and Michael Friebe. 2022. "Thyroid Nodule Region Estimation Using Auto-Regressive Modelling and Machine Learning." *Current Directions in Biomedical Engineering* 8 (2): 588–91. https://doi.org/10.1515/cdbme-2022-1150

Elmer Jeto Gomes Ataide*, Mathews S. Jabaraj, Alfredo Illanes, Simone Schenke, Axel Boese, Michael C. Kreissl, Michael Friebe

Thyroid Nodule Region Estimation using Auto-Regressive Modelling and Machine Learning

https://doi.org/10.1515/cdbme-2022-1150

Abstract: Ultrasound (US) imaging is used for the diagnosis and also evaluation of thyroid nodules. A Thyroid Imaging Reporting and Data System (TIRADS) is used for the risk stratification of thyroid nodules through US images. The composition of thyroid nodules plays an important role in the risk-stratification process. The percentages of cystic and solid components in a thyroid nodule are one of the features that are can be indicative of the risk of malignancy. In this work, we attempt to classify and estimate solid and cystic regions within nodules. 20x20 texture patches were extracted from solid and cystic regions and converted into signals. These signals are decomposed into low, mid, and high-frequency bands using Continuous Wavelet Transform (CWT). A total of 36 features were extracted from the decomposed signals using Auto-Regressive Modeling. The features were fed into three different Machine Learning (ML) algorithms (Artificial Neural Networks, K-Nearest Neighbors, and Random Forest Classifier) to provide us with a classification of solid versus cystic regions in thyroid nodule US images. The Random Forest Classifier obtained an Accuracy, Sensitivity, and Specificity of 90.41%, 99% and 91% respectively which was the highest among the three chosen ML algorithms. Additionally, the output from the classification phase was also be used to determine the percentage of cystic and solid regions with a given thyroid nodule US image.

***Elmer Jeto Gomes Ataide:** Otto-von-Guericke University Medical Faculty, Magdeburg, Germany, e-mail: elmer.gomesataide@ovgu.de
Matthews S. Jabaraj: Otto-von-Guericke University, Magdeburg, Germany
Alfredo Illanes: Surag Medical GmbH, Magdeburg, Germany.
Simone Schenke: Klinikum Bayreuth, Bayreuth Germany.
Axel Boese, Michael C. Kreissl: Otto-von-Guericke University Medical faculty, Magdeburg, Germany.
Michael Friebe: CEO IDTM GmbH, Recklinghausen, NRW, Germany. Professor, AGH UST, Department of Measurement and Electronics, Krakow, Poland

Keywords: Ultrasound Imaging, Thyroid Nodules, Feature Extraction, Region Estimation, Machine Learning, Classification.

1 Introduction

Thyroid nodules are mostly benign neoplasms, the incidence of which is higher in females compared to males [1]. Normal palpation of the neck results in a 4% to 7% chance of detecting a nodule [2, 3]. Ultrasound (US) imaging is used both, for the diagnosis and evaluation of thyroid nodules. Nodules are detected in up to 76% of the examined population when US is used for diagnosis [3 - 5].

Differentiating between cystic and solid regions in a nodule aids in the risk- stratification. Multiple studies have shown that the portion and percentages of cystic and solid components in a thyroid nodule can be indicative whether a nodule is benign or malignant [6, 7]. Physicians use three main types of standardized classification methods for the risk stratification of thyroid nodules. [8, 9, 10]. Each of these classification methods scores thyroid nodules with its own scoring system. For example, ACR –TIRADS [8] scores solid and cystic thyroid nodules with a more detailed approach as compared to KWAK-TIRADS [9]. Physicians experience high levels of subjectivity and inter observer variability while scoring thyroid nodules, even while using the same classification methods. This occurs particularly in the case of nodules with a mixed composition of solid and cystic regions. Thyroid nodules exhibiting a larger cystic portion are considered to be benign whereas those nodules that are predominantly solid have a higher risk of being malignant [9]. Therefore, it is helpful for the physician, to know these characteristics for a newly detected nodule and knowing their percentage is aimed at better risk-stratification. Additionally, the determination of the exact percentage of solid and cystic regions would aid the

physician in better calculation of doses for radioiodine therapy.

Automated assessment could help eliminate subjectivity significantly and hence improve the overall risk stratification of thyroid nodules. Other approaches for the classification of nodules have proven to be successful in terms of classifying using various approaches such as deep learning approaches, geometry and morphology and even local binary pattern variants [11 - 13]. But the drawbacks lie in the fact that these approaches use the whole image as an input. This means larger volumes of data are required and they do not isolate and consider thyroid compositions and regions individually. This makes the detailed risk stratification unreliable.

In this work we attempt to classify solid and cystic regions within thyroid nodules and aid physicians in objective thyroid nodule risk stratification. The core motivation behind this work is to provide physicians an objectively accurate region classification and estimation within the thyroid particularly with respect to solid and cystic regions in order to reduce subjectivity. The work done here is directed at solving a clinical need as expressed by the physicians. As a secondary motive, this work also aims at achieving comparable results with access to limited data.

2 Materials and Methods

2.1 Image and Texture Database

The database was collected at the Medical Faculty at Otto-von-Guericke University in Magdeburg in the Department of Nuclear Medicine. It consists of a total of 24 cases. Electromagnetically tracked US videos (12L probe and Logiq S8, GE Wisconsin, USA) of the thyroid were obtained. A total of 1019 US images containing thyroid nodules were extracted from the tracked videos. The number of images per case depended on the size number and orientation of the nodules present. The resolution of the images was 1442 x 899. The ground truth annotations for solid and cystic regions were done based on the input of four experienced physicians. A texture patch database of solid and cystic regions was generated based on this input. This was created in order to isolate, analyse and classify these two different regions within a thyroid nodule as well as compensate for limited data availability. Patch sizes of 20x20 pixels were extracted from the two selected regions independently. The patch size was determined based on previous experiments conducted as well as the literature [14]. Patches were selected without any overlap with each other. A total of 10190 solid texture patches were first extracted followed by 8980 cystic patches. Each of the patches was converted into signals following the methods put forth by [14]. Following this, CWT decomposition and AR modelling for feature extraction was carried out [14].

The approach proposed by [14] has proven to be effective in the classification and segmentation of the thyroid gland. In this work, we take into consideration the texture characterization properties and ability to bypass noise and US artefacts as shown by the work in [14]. The difference lies in the type frequency band decomposition and features extracted.

2.2 CWT decomposition and AR modelling for feature extraction

The signals were decomposed into three frequency bands, i.e. low, mid and high frequency bands. The low frequency bands were selected between the ranges of 8-24 Hz. The mid frequencies lay between 25-33 Hz while the high frequencies lay between 34-46 Hz. The CWT spectrum of each band was reconstructed into signals (low, mid and high) for further analysis and feature extraction using the Daubechies 4 (db4) wavelet. AR modelling was used to extract features from the bands. A total of 36 features (4 signal conversions X 3 decompositions X 3 main features = 36 features in total) were extracted that included Area under the Curve (AUC), Maximum Amplitude of the Curve (Max Amp) and Mean Value of the Curve (Mean C). The features were compiled for the 10190 signals and labelled as 0 and 1 (0 = cystic and 1 = solid). A final feature set of 19170 x 36 was generated and used in the classification phase

2.3 Classification

The classification of the texture patches was done by employing Artificial Neural Networks (ANN), K-Nearest Neighbours (KNN), and Random Forest Classifier. These classifiers were selected based on their different methods of classification as well as familiarity. Additionally, the different classification methods would help determine the most optimal algorithm for the inclusion in future work for the similar problem statement.

A multi- layer perceptron was used for the classification step coupled with backpropagation. The input layer consists of nine nodes which represent the features extracted. The output layers consist of two nodes which represent Solid and Cystic texture (labels 0 and 1). The network was built with 7 hidden layers

and the sigmoid function was used for activation. The learning rate was set for 0.001 and no dropout was used. [15 - 17]

We employed the Linear K-Nearest Neighbor algorithm as an Instance-based algorithm for classification. Minkowski distance algorithm to compute the similarity metric. In K-NN algorithm K is the number of neighbors induced in the examination [15, 18, and 19]. For this we used K= 5 which was determined to be the optimum number after experimenting with several other inputs for K between 1 and 10.

Decision trees from randomly selected subsets from the training set are created by Random forest Classifier with a total number of trees = 100. The votes from the various decision trees are aggregated to decide the final class of the object [18]. An 80-20 train-test split is used in this section. This were new images that are not part of the original training set and the testing set created by the internal train-test split. Additionally, Bispectrum analysis was used to also extract features from the same patch database and trained with the same models to act as a comparison.

3 Results and Discussion

Table 1 shows the classification results obtained using the three different algorithms for the features extracted using AR modelling. Table 2 shows the classification using the same algorithms but the features used for this were extracted using Bispectrum analysis. In Table 1 the ANN depicts an accuracy, sensitivity and specificity of 83.14%, 82% and 84% respectively. While in Table 2 using the bispectrum features it shows an accuracy, sensitivity and specificity of 78.29%, 79% and 80% respectively. The K-NN algorithm was able to classify the cystic and solid regions with an accuracy of 84.55%, sensitivity of 83% and Specificity of 86%. Similarly, in Table 2 using the bispectrum features K-NN shows an accuracy, sensitivity and specificity of 80.12%, 79% and 88% respectively. Among the three algorithms, Random Forest Classifier obtains the highest values with an accuracy of 90.41%, sensitivity of 99% and specificity of 91%. This is also holds true for the Random Forest results in Table 2 with an accuracy, sensitivity and specificity of 86.32%, 87% and 86%. All three algorithms are able to classify the cystic and solid regions with adequate accuracy in both tables. However, it should be noted that while comparing the AR modelled features to the bispectrum features, the AR modelled features result in a higher classification accuracy, sensitivity and specificity. Looking solely at the results from Table 1, the Random Forest Classifier outperforms the ANN and K-NN algorithms in the domains of sensitivity and specificity. This means that the Random Forest Classifier is able to more accurately determine what the cystic/solid region is and what is not better than the remaining two algorithms. This is why the Random Forest Classifier was chosen for generating the final image seen in Figure 1.

Table 1: Results of the Classification of solid and cystic regions in the thyroid nodules using four different algorithms using AR modelled features

Algorithm	Accuracy	Sensitivity	Specificity
ANN	83.14%	82%	84%
K-NN	84.55%	83%	86%
Random Forest	90.41%	99%	91%

Table 2: Results of the Classification of solid and cystic regions in the thyroid nodules using four different algorithms using Bispectrum modelled features.

Algorithm	Accuracy	Sensitivity	Specificity
ANN	78.29%	79%	80%
K-NN	80.12%	79%	88%
Random Forest	86.32%	87%	86%

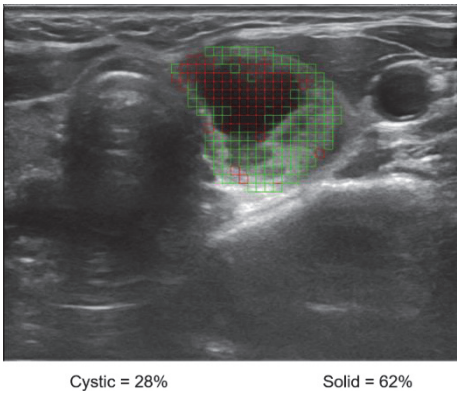


Figure 1: An example of the visual representation of the cystic (red) and solid (green) region estimation in terms of percentages obtained from the output of the classification phase

The output labels obtained from the unseen testing portion of classification phase of the Random Forest Classifier are further used to estimate the cystic and solid regions in terms of percentages for a given thyroid nodule in an US image. This is depicted in Fig. 1. From the figure can see that the features extracted can be used by classification algorithms to adequately detect and estimate in terms of percentages the cystic and solid regions within a thyroid nodule in a given US image.

4 Conclusion and Future Scope

In this work we attempted to classify and quantify solid and cystic regions in thyroid nodules with Machine Learning and while using a limited dataset. We do this by extracting patches from the available images to create a texture patch database followed by AR modelled features. The percentage of solid and cystic regions within a nodule can be obtained through these results. This would reduce subjectivity among physicians during the TIRADS classification. The next steps would be to improve the model performance by introducing texture patches from different US devices and work on a 3D visualization method for the physicians that provides them with accurate percentages of solid and cystic components within a thyroid nodule.

Author Statement

No funding involved. Authors state no conflict of interest. Informed consent has been obtained from all individuals included in this study. Ethical consideration was obtained under the study number RAD362-16/19.

References

[1] G. Popoveniuc and J. Jonklaas, “Thyroid Nodules,” Medical Clinics of North America, vol. 96, no. 2, pp. 329–349, 2012.

[2] P. A. Singer, “Treatment guidelines for patients with thyroid nodules and well-differentiated thyroid cancer. American Thyroid Association,” Archives of Internal Medicine, vol. 156, no. 19, pp. 2165–2172, 1996.

[3] J. F. Desforges and E. L. Mazzaferri, “Management of a Solitary Thyroid Nodule,” New England Journal of Medicine, vol. 328, no. 8, pp. 553–559, 1993.

[4] S. Ezzat, “Thyroid incidentalomas. Prevalence by palpation and ultrasonography,” Archives of Internal Medicine, vol. 154, no. 16, pp. 1838–1840, 1994.

[5] G. H. Tan, “Thyroid Incidentalomas: Management Approaches to Nonpalpable Nodules Discovered Incidentally on Thyroid Imaging,” Annals of Internal Medicine, vol. 126, no. 3, p. 226, Jan. 1997.

[6] D. Kim, E. Lee, H. In, and S. Kim, “Sonographic Differentiation of Partially Cystic Thyroid Nodules: A Prospective Study,” American Journal of Neuroradiology, vol. 31, no. 10, pp. 1961–1966, 2010.

[7] J. M. Park, Y. Choi, and H. J. Kwag, “Partially Cystic Thyroid Nodules: Ultrasound Findings of Malignancy,” Korean Journal of Radiology, vol. 13, no. 5, p. 530, 2012.

[8] F. N. Tessler, W. D. Middleton, E. G. Grant, J. K. Hoang, L. L. Berland, S. A. Teefey, J. J. Cronan, M. D. Beland, T. S. Desser, M. C. Frates, L. W. Hammers, U. M. Hamper, J. E. Langer, C. C. Reading, L. M. Scoutt, and A. T. Stavros, “ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee,” Journal of the American College of Radiology, vol. 14, no. 5, pp. 587–595, 2017.

[9] J. Y. Kwak, K. H. Han, J. H. Yoon, H. J. Moon, E. J. Son, S. H. Park, H. K. Jung, J. S. Choi, B. M. Kim, and E.-K. Kim, “Thyroid Imaging Reporting and Data System for US Features of Nodules: A Step in Establishing Better Stratification of Cancer Risk,” Radiology, vol. 260, no. 3, pp. 892–899, 2011.

[10] G. Russ, S. J. Bonnema, M. F. Erdogan, C. Durante, R. Ngu, and L. Leenhardt, “European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS,” European Thyroid Journal, vol. 6, no. 5, pp. 225–237, 2017.

[11] D. Koundal, “Computer-Aided Diagnosis of Thyroid Nodule: A Review,” International Journal of Computer Science & Engineering Survey, vol. 3, no. 4, pp. 67–83, 2012.

[12] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, and M. Eramian, “Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network,” Journal of Digital Imaging, vol. 30, no. 4, pp. 477–486, Oct. 2017.

[13] Gomes Ataide, Elmer Jeto, et al. “Thyroid Nodule Classification for Physician Decision Support Using Machine Learning-Evaluated Geometric and Morphological Features.” Sensors, vol. 20, no. 21, Oct. 2020, p. 6110., doi:10.3390/s20216110.

[14] A. Illanes, N. Esmaeili, P. Poudel, S. Balakrishnan, and M. Friebe, “Parametrical modelling for texture characterization—A novel approach applied to ultrasound thyroid segmentation,” Plos One, vol. 14, no. 1, 2019.

[15] N. Mahmoodian, A. Schaufler, A. Pashazadeh, A. Boese, M. Friebe. and A. Illanes, 2019. Proximal detection of guide wire perforation using feature extraction from bispectral audio signal analysis combined with machine learning. Computers in biology and medicine, 107, pp.10-17.

[16] J.M. Górriz, J.M., Ramírez, J., Puntonet, C.G. and Segura, J.C., 2006. An efficient bispectrum phase entropy-based algorithm for VAD. In Ninth International Conference on Spoken Language Processing.

[17] W.A. Zgallai, 2013. Second-and third-order statistical characterization of non-linearity and non-gaussianity of adult and fetal ECG signals and noise (Vol. 9). January.

[18] Y. Chen, Q. Wu, and X. He, 2011. Human action recognition based on radon transform. In Multimedia Analysis, Processing and Communications (pp. 369-389). Springer, Berlin, Heidelberg.

[19] F. Molinari, S.V. Sree, U.R. Acharya, J.S. Suri, S. Chattopadhyay, and Ng, K.H., 2011. Automated diagnosis of epileptic EEG using entropies.

[20] S. Chaplot, L.M. Patnaik, and N.R. Jagannathan, 2006. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomedical signal processing and control, 1(1), pp.86-92.

Publication 4

Gomes Ataide, Elmer Jeto, Nikhila Ponugoti, Alfredo Illanes, Simone Schenke, Michael Kreissl, and Michael Friebe. 2020. "Thyroid Nodule Classification for Physician Decision Support Using Machine Learning-Evaluated Geometric and Morphological Features." *Sensors* 20 (21). https://doi.org/10.3390/s20216110.

sensors

IMPACT FACTOR 3.9

Indexed in: PubMed

CITESCORE 6.8

Article

Thyroid Nodule Classification for Physician Decision Support Using Machine Learning-Evaluated Geometric and Morphological Features

Elmer Jeto Gomes Ataide, Nikhila Ponugoti, Alfredo Illanes, Simone Schenke, Michael Kreissl and Michael Friebe

Special Issue

Machine Learning for Biomedical Imaging and Sensing

Edited by

Dr. Andy Taylor and Dr. Jonine Figueroa

MDPI

https://doi.org/10.3390/s20216110

sensors

MDPI

Article

Thyroid Nodule Classification for Physician Decision Support Using Machine Learning-Evaluated Geometric and Morphological Features

Elmer Jeto Gomes Ataide [1,2,*], Nikhila Ponugoti [2], Alfredo Illanes [2], Simone Schenke [1], Michael Kreissl [1] and Michael Friebe [2,3]

1 Clinic for Radiology and Nuclear medicine, Department of Nuclear Medicine, Otto-von-Guericke University Medical Faculty, 39120 Magdeburg, Germany; simone.schenke@med.ovgu.de (S.S.); michael.kreissl@med.ovgu.de (M.K.)
2 INKA-Application Driven Research, Otto-von-Guericke University Magdeburg, 39120 Magdeburg, Germany; nikhila.ponugoti@gmail.com (N.P.); alfredo.illanes@ovgu.med.de (A.I.); michael.friebe@ovgu.de (M.F.)
3 IDTM GmbH, 45657 Recklinghausen, Germany
* Correspondence: elmer.gomesataide@ovgu.de

Received: 29 August 2020; Accepted: 26 October 2020; Published: 27 October 2020

Abstract: The classification of thyroid nodules using ultrasound (US) imaging is done using the Thyroid Imaging Reporting and Data System (TIRADS) guidelines that classify nodules based on visual and textural characteristics. These are composition, shape, size, echogenicity, calcifications, margins, and vascularity. This work aims to reduce subjectivity in the current diagnostic process by using geometric and morphological (G-M) features that represent the visual characteristics of thyroid nodules to provide physicians with decision support. A total of 27 G-M features were extracted from images obtained from an open-access US thyroid nodule image database. 11 significant features in accordance with TIRADS were selected from this global feature set. Each feature was labeled (0 = benign and 1 = malignant) and the performance of the selected features was evaluated using machine learning (ML). G-M features together with ML resulted in the classification of thyroid nodules with a high accuracy, sensitivity and specificity. The results obtained here were compared against state-of the-art methods and perform significantly well in comparison. Furthermore, this method can act as a computer aided diagnostic (CAD) system for physicians by providing them with a validation of the TIRADS visual characteristics used for the classification of thyroid nodules in US images.

Keywords: thyroid nodules; ultrasound imaging; TIRADS; feature extraction; machine learning; classification; computer aided diagnosis

1. Introduction

The thyroid is one of the largest endocrine glands located below the epiglottis. It is responsible for several physiological functions such as the production of hormones, regulation of brain and nerve cells, and development and functioning of organs like the heart, eyes, hair, skin, and intestines [1]. Irregularities and/or deformations of the thyroid lead to its inability to efficiently carry out these functions. Nodules within the thyroid are mostly benign neoplasms. Ultrasound (US) imaging is typically used as the first point of diagnosis and also for the evaluation of the thyroid nodules, as it effectively images and visualizes soft tissue structures. Additionally, it is free of ionizing radiation and is the most widely available and affordable imaging modality [2].

The assessment and evaluation of thyroid nodules using US imaging is done by the physician based on the visual characteristics observed in the scan. For this purpose, the Thyroid Imaging Reporting

and Data System (TIRADS) approach is used for the risk stratification and classification of benign vs. malignant nodules. Multiple versions of TIRADS exist such as that of the American Council of Radiology (ACR) TIRADS [3], Kwak-TIRADS [4], etc. Each of these versions differs, but all consider visual and textural features, such as margin, shape, calcification, composition, size, echogenicity and offer a scoring scheme that enables the physician to assess nodule malignancy.

The use of TIRADS has helped to standardize the evaluation of thyroid nodules found in US images. However, there still exists a considerable amount of inter-observer variability and overall subjectivity. To address the issue of subjective diagnoses several computer aided diagnostic (CAD) methods were proposed. These CAD methods use various feature extraction and classification algorithms to characterize thyroid nodules using US images into benign and malignant.

The aim of our work was to develop a more objective diagnostic approach for thyroid nodules using US images. The related studies significant to our work will be discussed in the following sub-section.

Related Work

Several methods have been proposed for feature extraction and computer-aided classification of thyroid nodules using US images. Apart from texture-based features [5], there are other types of feature extraction methods that can be used, like general shape-based feature extraction/classification using different image modalities, as well as shape-based feature extraction for the classification of thyroid nodules in US images. For that we also considered studies that have used the same database that we accessed.

Jianhua Liu et al. introduced the use of shape features of an image. The boundary and region-based feature extraction methods are explained in [6]. The use of shape features was observed in several applications for medical image analysis. Riti et al. employed shape features for the classification of lung cancer from computed tomography (CT) images and obtained an overall classification accuracy of 85% [7]. The same was seen in the work of Ferreira Junior et.al. where the margin sharpness was used for the classification of lung nodules in CT images [8]. Hiremath et al. suggested a shape feature-based approach for detecting follicles in ovarian ultrasound images. The follicles were classified depending on medical knowledge on parametrically defined measures, such as area, compactness, centroid, etc. [9]. Huang et al. in their research suggested the use of functional morphological characteristics to differentiate between benign and malignant breast tumors efficiently. Nineteen morphological features were extracted from ultrasound images and used for the classification of tumors. They obtained a classification accuracy of 82% and a sensitivity of 94% using a support vector machine (SVM) classifier [10]. Nugroho et al. also made use of shape-based feature analysis and extraction for the classification of breast nodules with a specific focus on the marginal characteristics of uncircumscribed versus circumscribed margins [11].

In the case of thyroid nodules, Gopinathan et al. performed thyroid nodule risk stratification and classification by analyzing the roundness and irregularity of nodule margins while using US and fine-needle biopsy [12]. Zulfanahri et al. suggested a system that can classify thyroid nodules in two groups, i.e., round to oval and irregular shapes using three chosen characteristics. The suggested system achieved an accuracy of 91.52% and specificity of 91.35% [13]. Similarly, Lina Choridah et al. proposed a technique to classify thyroid nodules based on marginal features. The suggested strategy effectively classified the thyroid nodules into two smooth and uneven classes using US images and obtained an accuracy and sensitivity of 92.30% and 91.88% respectively [14]. The image pattern classification technique used by Junying Chen et al. to categorize benign and malignant thyroid nodules proved efficient in the classification process to verify the types of pattern properties that could be used for the classification of thyroid nodules in US images [15].

Ding et al. defined statistical characteristics and texture based on elastography of the thyroid lesion area. The selection of features was then done using the algorithm called minimum redundancy -maximum-relevance. The selected features were then plugged into an SVM classifier [16].

Isa et al. used different multi-layer perceptron (MLP) models for the detection of thyroid diseases [17]. Statistical features coupled with demographic details of a sample were fed into three different classification algorithms such as random forest classifier (RFC), SVM, and logistic regression by Patricio et al. to differentiate between thyroid nodules [18]. Song et al. in their work with thyroid nodule US image classification proposed a hybrid multi-branch convolutional neural network (MBCNN) based on a feature cropping method for feature extraction. This work used an open-access dataset [19] as well as a local dataset and obtained a classification accuracy of 96.13% [20]. Chi et al. by using a fine-tuned deep convolutional neural network (FDCNN) approach obtained a classification accuracy of 98.29% [21]. Using the same dataset, Koundal et al. proposed a complete image-driven thyroid nodule detection approach that was able to detect thyroid nodules with an accuracy of up to 93.88% [22]. Nanda and Martha in their work with cancerous thyroid nodule detection employed local binary pattern variants (LBPV)-based feature extraction to classify between benign and malignant thyroid nodules with an accuracy of 94.5% [23] using [19].

In the presented literature it can be seen that there are several methods for the feature extraction and classification of thyroid nodules. Studies [6–11] give a broad understanding of the use of shape-based features in medical image analyses. Using various combinations of these shape-based features, studies [12–15] were able to classify thyroid nodules in US images with significant outcomes. Deep learning-based methods [20,21] result in an automatic feature extraction and classification method that differs from handcrafted features. The methods seen in the literature perform well, but most of them do not take into account the geometrical and morphological attributes of thyroid nodules following TIRADS. Since physicians exhibit significant levels of trust in TIRADS, it is essential to consider the features put forth by [3,4]. These are the attributes that are visible to the physicians on which their decision is partly based. Even the studies [12–15] that do take into account shape features do so only to a limited extent.

The purpose of this study is to use geometric and morphological feature extraction that takes into consideration features that are closely related to the visual shape-based (TIRADS) features currently used by physicians. This provides them with additional information and mathematical evidence to support their current TIRADS-based classification with an extra layer of objectivity.

In this work, we mainly focus on shape-based geometric and morphological feature extraction for the classification of thyroid nodules as either benign or malignant. Examining physicians use visual and textural characteristics to classify a nodule. Geometric and morphological features represent the visual aspect. The performance of the features extracted was evaluated using a RFC. The results obtained from the RFC were compared against other feature extraction and classification techniques found in the state of the art that use the same database. Furthermore, our approach was also compared to other shape-based feature extraction and classification methods found in the literature. The rest of the paper is structured as follows: Section 2 details the materials and methods used in this work. Section 3 presents the results and comparisons drawn with other feature extraction and classification methods, and the final Section 4 then provides a conclusion and discusses the next steps.

2. Materials and Methods

This section is divided into three parts: (1) database and data augmentation, (2) feature extraction and selection, and (3) classification. The section on the database and data augmentation gives details of the dataset used and the various data augmentation techniques used to balance and maximize the data usage. The second section presents details on the geometrical and morphological features extracted from the dataset and discusses the feature selection methods tested and employed for the selection of the most optimal features. The classification section highlights the classification method used to evaluate the selected features with subsequent classification into benign and malignant. The workflow diagram consisting of each step is presented in Figure 1.

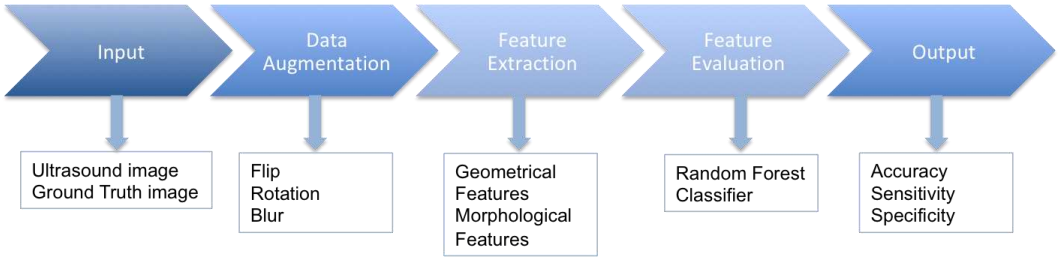


Figure 1. Workflow diagram.

2.1. Dataset and Data Augmentation

For this study, we used the Digital Database of Thyroid Ultrasound Images (DDTI) open-access dataset of thyroid nodule ultrasound images from the Instituto de Diagnostico Medico (IDIME) [19]. The dataset consists of a total of 99 cases with 2D ultrasound images from different patients that are annotated and classified based on TIRADS classification. The dataset is divided into JPEG files and XML files. Each image has a resolution of 560 × 315. Each image has a corresponding XML file. The XML files provided a detailed classification for each of the nodules. The ground truth (GT) for each of the nodules was generated by experienced physicians and is available in the form of coordinates in the XML file. An example of benign and malignant nodules along with their GTs can be seen in Figure 2. In this study, we only considered two labels (benign and malignant) rather than all the TIRADS labels.

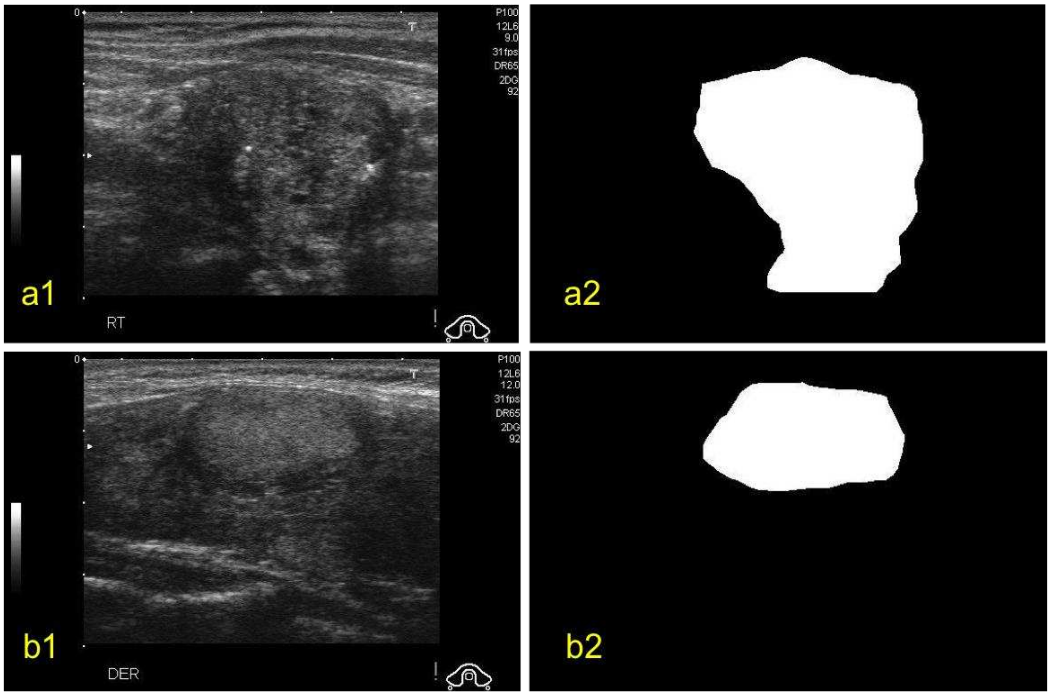


Figure 2. Examples of ultrasound images of thyroid nodules (**a1**) malignant nodule, (**a2**) its ground truth, (**b1**) benign nodule and (**b2**) its ground truth [19].

These cases were divided into 17 benign cases and 82 malignant cases. To correct the data imbalance, data augmentation techniques such as flipping, rotation and blurring were employed. Data augmentation was first used to balance the data and then again to further augment it. Finally, a total of 3188 images were obtained (1594 Malignant + 1594 Benign) from the original 134 images and used in the feature extraction phase of the study. The data augmentation was performed using the Augmentor python package. The percentages of each of the operations (flipping, rotation and blurring) cannot be estimated as they are applied stochastically.

2.2. Feature Extraction

In this work, two types of feature were extracted from the GTs of the US thyroid images. These were in the form of geometric and morphological (G-M) features. A summary of all the extracted features can be seen in Table 1. Figure 3 gives a visual depiction of a few geometric and morphological features.

Table 1. Overview of 27 extracted geometric and morphological (G-M) features.

Sr. No.	Features	Type
1	Convex Hull	**Geometric Features**
2	Convexity	
3	Solidity	
4	Elongation	
5	Compactness	
6	Rectangularity	
7	Orientation	
8	Roundness	
9	Major Axis Length	
10	Minor Axis Length	
11	Eccentricity	
12	Circular Variance	
13	Elliptic Variance	
14	Ratio of Major Axis Length to Minor Axis Length	
15	Bounding Box	
16	Centroid	
17	Convex Area	
18	Filled Area	
19	Convex Perimeter	
20	Area	**Morphological Features**
21	Perimeter	
22	Aspect Ratio	
23	Area Perimeter (AP)Ratio	
24	Object Perimeter to Ellipse Perimeter (TEP) Ratio	
25	TEP Difference	
26	Object Perimeter to Circular Perimeter (TCP) Ratio	
27	TCP Difference	

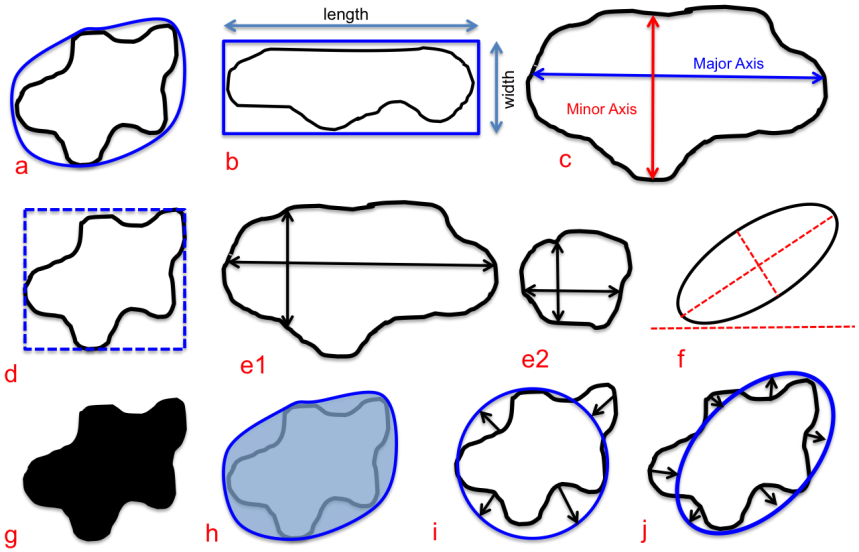


Figure 3. Visual depiction of geometric and morphological features. (**a**) Convexity, (**b**) elongation, (**c**) major and minor axes, (**d**) bounding box, (**e1**,**e2**) different instances of eccentricity, (**f**) orientation, (**g**) filled area, (**h**) convex area, (**i**) circular variance and (**j**) elliptical variance.

2.2.1. Geometric Features

Geometric features are those features that are used to construct an object with certain geometric elements such as points, curves, and lines as well as information related to edges that describe the shape or irregularity of a given boundary [24,25]. Note that all instances of the mention of the word "object" refer to thyroid nodules. The geometric features extracted from the images are as follows.

Convex Hull:

The convex hull of an object is the smallest convex structure within which an object is positioned. It is the smallest convex polygon that can contain the object. [26,27]:

Convexity:

Convexity is a function that measures the ratio of the convex hull with respect to the original contour of the shape. In this case, a convex hull is drawn around the original contour of the object. Convexity is calculated using the following equation [28]:

$$\text{Convexity} = \frac{P_c}{P_n} \tag{1}$$

where P_c = perimeter of the convex hull and P_n = perimeter of the object

Solidity:

Solidity helps describe the extent of a shape's convexity or concavity. It is given by Equation [29]:

$$\text{Solidity} = \frac{A_n}{A_c} \tag{2}$$

where A_n = area of the object and A_c = area of the convex hull. A value of a solid object is 1 and an object with an irregular boundary is defined with a value less than 1.

Elongation:

Elongation is the feature that measures the ratio between the length and width of a bounding box around an object. The result is a nodule's elongation measurement given as a value of 0 to 1. The object is approximately square or round shaped when the ratio is equal to 1. When the ratio is lower than 1, the object is more elongated. The equation can be seen below [26,30]:

$$\text{Elongation} = \frac{W_n}{L_n} \tag{3}$$

where W_n = width of the object and L_n = length of the object

Compactness:

Compactness is the ratio between the areas of an object with respect to the area of a circle with a perimeter equal to that of the object. It is given by the equation below [26,30]:

$$\text{Compactness} = \frac{4\pi A_n}{P_n^2} \tag{4}$$

where P_n = perimeter of the object and A_n = Area of the object

Rectangularity:

Rectangularity is defined as the ratio between the object area and the area of the minimum-bounding rectangle [31]. When an object returns a rectangularity of 1 it is said to be a rectangular object.

$$\text{Rectangularity} = \frac{A_n}{A_r} \tag{5}$$

where A_n = area of the object and A_r = Area of the rectangle.

Roundness:

The roundness of an object is defined as the ratio between the area of the object and the area of a circle with the same convex perimeter. It can be represented by the following equation [26,32]:

$$\text{Roundness} = \frac{4\pi A_n}{P_c^2} \tag{6}$$

where P_c = the convex perimeter of the object and A_n = the area of the object

Major Axis length:

The major axis of an object is the endpoints (X, Y) of the longest line traced through the object. The endpoints of the major axis ($X1$, $Y1$, and $X2$, $Y2$) are determined by calculating the pixel distance between every combination of border pixels in the object boundary. This is used to find the pair with the maximum length. The object's major axis length is the pixel distance between the major axis endpoints and is defined by the equation [26,32]:

$$\text{Major-axis length} = \sqrt{(X2 - X1)^2 + (Y2 - Y1)^2} \tag{7}$$

Minor Axis length:

The minor axis is the (x, y) endpoints of the longest line drawn by the object while still perpendicular to the major axis. The endpoints ($x1$, $y1$, and $x2$, $y2$) of the minor axis are calculated by computing the pixel distance between the two border pixel endpoints. The minor-axis length of an object is the pixel distance between the minor axis endpoints and is defined by the equation [26,32]:

$$\text{Minor-axis length} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \tag{8}$$

Eccentricity:

Eccentricity is defined as the ratio between minor axis length and the major axis length of the object [26]. The result is a measure of the eccentricity of the object, given as a value from 0 to 1.

$$\text{Eccentricity} = \frac{L_{ma}}{L_{MA}} \tag{9}$$

where L_{ma} = length of minor axis and L_{MA} = Length of major axis

Circular Variance:

Circular variance (C_{va}) is the comparison of an object's shape with respect to a known shape such as a circle. The circular variance is the object's proportional mean-squared error with respect to the solid circle [26,32]. This returns a null value for a perfect circle and increases and the shape and complexity change. It is given by the following equation.

$$C_{va} = \frac{\sigma R}{\mu R} \tag{10}$$

where μR and σR are the mean and standard deviation of the radial distance from the centroid (c_x, c_y) of the shape to the boundary points (x_i, y_i), $i \in [0, N-1]$.

They are represented by the formulae:

$$\mu R = \frac{1}{N}\sum_{i=1}^{N-1} d_i \tag{11}$$

and

$$\sigma R = \sqrt{\frac{1}{N}\sum_{i=1}^{N-1}(d_i - \mu R)^2} \tag{12}$$

where $d_i = \sqrt{(x_i - c_x)^2 + \left(y_i - c_y\right)^2}$

Elliptic Variance:

Elliptic variance (E_{va}) is measured similarly to the circular variance. An oval is fitted to the shape (rather than a circle), and the mean squared error is estimated [32].

$$E_{va} = \frac{\sigma' R}{\mu' R} \tag{13}$$

$$\mu' R = \frac{1}{N} \sum_{i=1}^{N-1} d'_i \tag{14}$$

and

$$\sigma' R = \sqrt{\frac{1}{N} \sum_{i=1}^{N-1} (d'_i - \mu' R)^2} \tag{15}$$

where

$d'_i = \sqrt{V_i^T \times C_{ellipse}^{-1} \times V_i}$

$C_{ellipse}^{-1}$ = Inverse of covariance matrix of the shape (ellipse)

$V_i = \left(\frac{x_i - c_x}{y_i - c_y}\right)$ and V_i^T = Transpose of V_i

Ratio of Major Axis Length to Minor Axis Length:

Thus is the ratio between major axis lengths to minor axis length.

$$\text{Ratio of length of Major axis and minor axis } = \frac{L_{MA}}{L_{ma}} \tag{16}$$

where L_{MA} = length of major axis and L_{ma} = length of minor axis.

Orientation:

The orientation is angle between the x-axis and the major axis of the object. It can also be defined as the direction of the shape [26].

Bounding Box:

The bounding box is the region's smallest rectangle that envelops the object [25]. Dimensions for the bounding box are those equal to the major and the minor axes.

$$\text{Area of bounding box } = L_{MA} \times L_{ma} \tag{17}$$

where L_{MA} = length of major axis and L_{ma} = length of minor axis.

Centroid:

The centroid is defined as the center of gravity of the object [25].

Convex Area:

The convex area of a nodule is the area surrounded by the convex hull [31].

Filled Area:

Is the total number of pixels within the marked object mask. [26]

Convex Perimeter:

The convex perimeter of an object is the perimeter of the convex hull that encloses the object [25].

2.2.2. Morphological Features

Morphological features are those features that define an object's structuring elements such as area, perimeter, aspect ratio, etc. [33]. The following morphological features were considered for the classification of thyroid nodules in this work.

Area:

Area is the space occupied by objects on a plane surface. Here area is defined as the number of pixels inside the object region [34].

Perimeter:

The number of pixels within the object border is the perimeter [26,34]. If $x1 \ldots xN$ is a list of boundaries, and the perimeter is defined by:

$$\text{Perimeter} = \sum_{i=1}^{N-1} di = \sum_{i=1}^{N-1} \left| X_i - X_{i+1} \right| \tag{18}$$

Aspect Ratio:

The aspect ratio is defined as the ratio between the tumor's depth and width [26]:

$$\text{Aspect Ratio} = \frac{D_n}{W_n} \tag{19}$$

where D_n = Depth of the object and W_n = Width of the object.

AP Ratio (area to perimeter (AP) ratio):

The AP ratio is the ratio between object area and perimeter of the object, and it is defined as:

$$\text{AP Ratio} = \frac{A_n}{P_n} \tag{20}$$

where A_n = Area of the object and P_n = Perimeter of the object

TEP Ratio (object perimeter to ellipse perimeter ratio):

The TEP ratio is the ratio of perimeters of an object to the related ellipse [17].

$$\text{TEP Ratio} = \frac{P_n}{P_e} \tag{21}$$

where P_n = perimeter of the object and P_e = perimeter of the ellipse

TEP Difference:

TEP is determined by the difference between the object perimeter and the related ellipse [17]:

TCP Ratio (object perimeter to circle perimeter ratio):

The TCP ratio is the ratio of the perimeter of the object to the relevant circle [17]:

$$\text{TCP Ratio} = \frac{P_n}{P_c} \tag{22}$$

where P_n = perimeter of the object and P_c = perimeter of the circle.

TCP Difference:

The TCP difference is known as the difference between the perimeter of the object and the corresponding circle [17].

2.2.3. Feature Selection

Feature selection was based on the most relevant visual characteristics in accordance with TIRADS. This was determined with the help of the two physicians in the Department of Nuclear Medicine at the University Hospital in Magdeburg, Germany. Clinicians use TIRADS classification that classifies a nodule based on its geometrical attributes such as shape, size, irregularity in margins and orientation. The 11 selected features provide the closest estimation to these attributes used by TIRADS. In order to further validate this claim the authors tested the performance of all the geometric and morphological features (Global) as well as those that were not selected (discounted). The performance metrics for each of these are seen in Table 4 in Section 3.

Labels were added to each row of features depending on whether they belonged to 0 = benign or 1 = malignant. An overview of all the extracted features can be seen in Table 1. The feature extraction

process was carried out using MATLAB 2018b. The feature sets were then exported as .csv files and used for the classification process. The performance of the features was evaluated using a RFC.

2.3. Classification

An RFC was trained for a binary classification problem where each of the feature rows was labeled as 0 or 1. RFC is a type of ensemble learning that builds a final classifier by using weak individual classifiers i.e., binary decision trees. Each tree is a collection of nodes and features that lead to the final classification result. The aggregation of the results from each of the individual trees is considered the final result of the classifier. The features were used as the independent variables and the labels as the dependent variables. The parameters chosen for the RFC are shown in Table 2. A train-test split of 70–30% was used on the data. The classification using RFC was undertaken in Python 3.7 using the scikit-learn library.

Table 2. Selected random forest classifier (RFC) parameters.

Parameter	Value
Number of Decision Trees	400
Criterion	Entropy
Bootstrap	True

3. Results and Discussion

A total of 27 features were extracted from the thyroid nodule dataset. The feature selection step led to the selection of the 11 most significant features. We selected three prime metrics to compare the G-M features against the performance of global and discounted features in our study for the feature selection as well as the methods found in the literature. The metrics selected were accuracy, sensitivity, and specificity.

The 11 significant features are given in Table 3. These features were selected based on clinical input and expertise. Further experiments confirmed the validity of the features selected. This is highlighted in Table 4.

Table 3. Eleven most significant features selected from the global feature list of 27.

Sr. No.	Features	Type
1	Solidity	**Geometric Features**
2	Orientation	
3	Roundness	
4	Major Axis Length	
5	Minor Axis Length	
6	Bounding Box	
7	Convex Area	
8	Area	**Morphological Features**
9	Perimeter	
10	Aspect Ratio	
11	AP Ratio	

Table 4. Performance metrics of selected features versus global and discounted features.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
Global	70.18	48.07	92.29
Discounted	61.55	31.65	91.45
G-M	99.33	99.39	99.25

As can be seen the accuracy scores of global features and discounted features are considerably lower (70.18% and 61.55% respectively) than compared to G-M. Additionally, both global and discounted features exhibit high specificities, but the sensitivities were low (48.07% and 31.65% respectively). As this is a case of cancer classification, more focus was given to the sensitivity score due to its clinical relevance (true positive rates). The selected 11 final features resulted in high accuracy, specificity and sensitivity scores. i.e., the classifier was able to identify benignity and malignancy of the nodules much more effectively.

The selected 11 features were fed into the RFC and the results obtained were compared against the results from [20,21,23] that used deep learning-based and LBPV-based methods. These studies used feature extraction and classification techniques that were different from our proposed approach but were all tested on the same open-access dataset [19]. Additionally, the results were further compared against other shape-based feature extraction techniques [13,14] found in the literature.

The same performance metrics were used to present the results obtained from the classifiers. This is presented in Tables 5 and 6 below. The comparison is done in two steps. The first step involves the comparison between the proposed method and the methods found in the state of the art that used the same dataset as us. The second step draws a comparison between the proposed method and other thyroid nodule feature extraction and classification approaches using shape-based features found in other related studies.

Table 5. Feature evaluation using RFC compared to the performance of the related approaches using the same dataset.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
MBCNN [20]	96.13	97.18	-
FDCNN [21]	98.29	99.10	93.90
LBPV (SVM) [23]	94.5	97.25	94.50
G-M (RFC)	99.33	99.39	99.25

Table 6. Feature evaluation using RFC compared to the performance of shape-based features found in related studies using different datasets.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
Margin Features [13]	91.52	91.80	91.35
Margin Features [14]	92.30	91.88	92.73
G-M	99.33	99.39	99.25

Table 5 shows the classification results obtained from our proposed feature extraction approach, G-M and three different thyroid nodule feature extraction and classification approaches found in the state of the art. i.e., MBCNN [20], FDCNN [21] and LBPV [23]. Each of these methods uses a different feature extraction and classification method from the proposed approach. However, the dataset used in all four cases is the same [19]. Data augmentation was used by [20,21]. It can be seen from the table that G-M based features display a higher accuracy of 99.33% as compared to [20,21,23] while using the same database. G-M features also result in a high specificity score of 99.25%. In the case of sensitivity, G-M features exhibit a score of 99.39%. From the depicted results it can be inferred that the proposed feature extraction method can classify benign and malignant thyroid nodules with high accuracy. Additionally, this approach also displays high true positive (Sensitivity) and true negative (Specificity) rates. This means that the feature extracted can be distinguished well.

Table 6 shows the classification results obtained from the proposed G-M feature extraction method compared to two approaches found in the literature that use shape-based features for the classification of thyroid nodules. Each of the two studies [13,14] compared to G-M uses features that help characterize the nodules based on the extent of the margins being oval. Both the compared methods use different datasets in their work. It can be seen from the table that the G-M features when fed into the RFC

classifier exhibit an accuracy, sensitivity, and specificity of 99.33%, 99.39%, and 99.25%, respectively. This is a considerable improvement over the same metrics seen in [13,14].

In both comparison cases, across different feature extraction methods using the same dataset as well as similar feature extraction methods using different input data, the G-M approach outperforms the state-of-the-art. Even though [13,14,20,21,23] show high accuracies, sensitivities, and specificities, the features extracted are often not relevant for the physician performing the examination. To a certain extent, [13,14] take into account some shape features. But, [20,21,23] use features that are not in accordance with TIRADS. While reviewing US images of thyroid nodules physicians take into consideration the visual and textural characteristics of a nodule to classify it as benign or malignant. The geometric and morphological features encompass the visual characteristics of a nodule. These are the closest estimation to the features defined by the gold standard in TIRADS. This can be observed across all three calculated metrics. It is evident that the G-M features adequately emulate the visual characteristics that are defined in TIRADS. These visual aspects of a thyroid nodule such as margins irregularity and shape can be directly attributed to the G-M features extracted in this work.

4. Conclusions and Future Work

This study focused on the geometric and morphological feature extraction techniques for the classification of thyroid nodules from US images. This work used a total of 3188 images. A total of 27 geometric and morphological features were defined and extracted from these images and then the 11 most significant features selected in accordance with the TIRADS-based visual features and labeled based on their class (0 = benign, 1 = malignant). The performance of the selected features was then evaluated by the classification accuracy, sensitivity, and specificity obtained from the RFC.

The consideration of G-M features for a computer-aided diagnostic approach for thyroid nodules is clinically relevant. The 11 selected features in this study proved to be the best combination from the overall feature set of 27 because these selected features provide information such as the shape, irregularity in the boundary, orientation, and size of the US thyroid nodule. However, it must be noted that G-M features are only one part of the features that help in the classification. According to TIRADS, physicians also need to consider texture features found in a nodule, which were not considered for this work. Hence, a notable observation of the proposed approach is that it only considers one classification aspect. When G-M features are combined with texture features the overall accuracy might change, which needs to be studied further. Another limitation of this work is that everything was carried out on a single open-source database and further validation of the approach needs to be carried using additional datasets including different ultrasound scanners. We are currently working on acquiring data at the Department of Nuclear Medicine at the University Hospital in Magdeburg, Germany. This would help to improve the robustness of the features extracted across different datasets.

In future work, we would like to omit the use of data augmentation techniques all together. However, this is dependent on the amount of data that is being currently collected as mentioned. Until large volumes of US thyroid nodule images are available, there are two possible strategies that can be employed additionally to test the relevance of the selected features. The first is to augment the data while being aware of the percentage of each operation (flipping, rotation and blurring) performed. This would help us determine and understand how each of the selected features behaves with respect to each data augmentation operation. Additionally, it would also help us determine the extent to which the data augmentation effects the final classification. The second method would be to test the features extracted from non-augmented images against features extracted from augmented ones and to show the deviation between these two sets.

The addition of texture-based features to the G-M features would provide a larger feature set that would then consider both aspects of a TIRADS classification and provide physicians with better decision support for the classification of thyroid nodules. Hence, a step towards reducing inter-observer variability and overall diagnostic subjectivity.

Author Contributions: Conceptualization, E.J.G.A., A.I., S.S., M.K. and M.F.; methodology, E.J.G.A., A.I. and S.S.; software, E.J.G.A. and N.P.; validation, E.J.G.A., A.I. and S.S.; formal analysis, E.J.G.A., N.P. and A.I.; investigation, E.J.G.A. and S.S.; writing—original draft preparation, E.J.G.A.; writing—review and editing, A.I., S.S., M.K. and M.F.; supervision, M.K. and M.F.; project administration, E.J.G.A.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. InformedHealth.org. How Does the Thyroid Gland Work? Available online: https://www.ncbi.nlm.nih.gov/books/NBK279388/ (accessed on 26 June 2019).
2. Thomas, L.S. *Diagnostic Ultrasound Imaging: Inside Out*; Elsevier Academic Series: New York, NY, USA, 2004; Chapter 1.
3. Tessler, F.N.; Middleton, W.D.; Grant, E.G.; Hoang, J.K.; Berland, L.L.; Teefey, S.A.; Cronan, J.J.; Beland, M.D.; Desser, T.S.; Frates, M.C.; et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J. Am. Coll. Radiol.* **2017**, *14*, 587–595. [CrossRef] [PubMed]
4. 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*, 892–899. [CrossRef]
5. 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? *Eur. J. Radiol.* **2018**, *99*, 1–8. [CrossRef] [PubMed]
6. Liu1, J.; Shi, Y. *Image Feature Extraction Method Based on Shape Characteristics and Its Application in Medical Image Analysis*; Department of Software Institute, Department of Information Engineering, North China University of Water Resources and Electric Power: Zhengzhou, Henan Province, China, 2011.
7. Riti, Y.F.; Nugroho, H.A.; Wibirama, S.; Windarta, B.; Choridah, L. Feature extraction for lesion margin characteristic classification from CT Scan lungs image. In Proceedings of the 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 23–24 August 2016; pp. 54–58.
8. Junior, J.R.F.; Oliveira, M.C. Evaluating Margin Sharpness Analysis on Similar Pulmonary Nodule Retrieval. In Proceedings of the 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, Sao Paulo, Brazil, 22–25 June 2015; pp. 60–65.
9. Hiremath, P.S.; Tegnoor, J.R. *Recognition of Follicles in Ultrasound Images of Ovaries Using Geometric Features*; Department of Studies and Research in Computer Science, Gulbarga University: Gulbarga, Karnataka, India, 2009.
10. Huang, Y.L.; Chen, D.R.; Jiang, Y.R.; Kuo, S.J.; Wu, H.K.; Moon, W.K. *Computer-Aided Diagnosis Using Morphological Features for Classifying Breast Lesions on Ultrasound*; Wiley InterScience: Hoboken, NJ, USA, 2008.
11. Nugroho, H.A.; Triyani, Y.; Rahmawaty, M.; Ardiyanto, I. Analysis of margin sharpness for breast nodule classification on ultrasound images. In Proceedings of the 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), Phuket, Thailand, 12–13 October 2017; pp. 1–5.
12. Anil, G.; Hegde, A.; Chong, F.H.V. Thyroid nodules: Risk stratification for malignancy with ultrasound and guided biopsy. *Cancer Imaging* **2011**, *11*, 209. [PubMed]
13. Zulfanahri; Nugroho, H.A.; Nugroho, A.; Frannita, E.L.; Ardiyanto, I. Classification of Thyroid Ultrasound Images Based on Shape Features Analysis. In Proceedings of the 2017 Biomedical Engineering International Conference, Yogyakarta, Indonesia, 6–7 November 2017.
14. Nugroho, H.A.; Frannita, E.L.; Nugroho, A.; Zulfanahri; Ardiyanto, I.; Choridah, L. Classification of thyroid nodules based on analysis of margin characteristic. In Proceedings of the 2017 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Jakarta, Indonesia, 23–26 October 2017; pp. 47–51. [CrossRef]
15. Chen, J.; You, H. Efficient classification of benign and malignant thyroid tumors based on characteristics of medical ultrasonic images. In Proceedings of the 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, Shanxi, China, 3–5 October 2016; pp. 950–954.

16. Ding, J.; Cheng, H.; Huang, J.; Zhang, Y.; Ning, C. A novel quantitative measurement for thyroid cancer detection based on elastography. In Proceedings of the 2011 4th International Congress on Image and Signal Processing (CISP), Shanghai, China, 15–17 October 2011; pp. 1801–1804.
17. Isa, I.S.; Saad, Z.; Omar, S.; Osman, M.; Ahmad, K.; Sakim, H.M. Suitable MLP network activation functions for breast cancer and thyroid disease detection. In Proceedings of the 2010 Second International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM), Bali, Indonesia, 28–30 September 2010; pp. 39–44.
18. Patrício, M.; Oliveira, C.; Caseiro-Alves, F. Differentiating malignant thyroid nodule with statistical classifiers based on demographic and ultrasound features. In Proceedings of the 2017 IEEE 5th Portuguese Meeting on bioengineering (ENBENG), Coimbra, Portugal, 16–18 February 2017; pp. 1–4.
19. Pedraza, L.; Vargas, C.; Narváez, F.; Durán, O.; Muñoz, E.; Romero, E. An open-access thyroid ultrasound-image Database. In Proceedings of the Conference: 10th International Symposium on Medical Information Processing and Analysis, Cartagena de Indias, Colombia, 14–16 October 2014.
20. Song, R.; Zhang, L.; Zhu, C.; Liu, J.; Yang, J.; Zhang, T. Thyroid Nodule Ultrasound Image Classification Through Hybrid Feature Cropping Network. *IEEE Access* **2020**, *8*, 64064–64074. [CrossRef]
21. Chi, J.; Walia, E.; Babyn, P.; Wang, J.; Groot, G.; Eramian, M. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. *J. Digit. Imaging* **2017**, *30*, 477–486. [CrossRef] [PubMed]
22. Koundal, D.; Gupta, S.; Singh, S. Computer aided thyroid nodule detection system using medical ultrasound images. *Biomed. Signal Process. Control* **2018**, *40*, 117–130. [CrossRef]
23. Nanda, S.; Mar, M. Identification of Thyroid Cancerous Nodule using Local Binary Pattern Variants in Ultrasound Images. *Int. J. Eng. Trends Technol.* **2017**, *49*, 369–374. [CrossRef]
24. Mortenson, M.E. *Geometric Modeling*; John Wiley: New York, NY, USA, 1985; p. 763.
25. Yang, M.; Kpalma, K.; Ronsin, J. A Survey of Shape Feature Extraction Techniques. *Peng Yeng Yin Pattern Recognit.* **2008**, *15*, 43–90.
26. Sagar, C. Feature selection techniques with R. *Data Aspirant*, 15 January 2018.
27. Moltz, J.; Bornemann, L.; Kuhnigk, J.-M.; Dicken, V.; Peitgen, E.; Meier, S.; Bolte, H.; Fabel, M.; Bauknecht, H.; Hittinger, M.; et al. Advanced Segmentation Techniques for Lung Nodules, Liver Metastases, and Enlarged Lymph Nodes in CT Scans. *IEEE J. Sel. Topics Signal Process.* **2009**, *3*. [CrossRef]
28. Kubota, T.; Jerebko, A.K.; Dewan, M.; Salganicoff, M.; Krishnan, A. Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models. *Med. Image Anal.* **2011**, *15*, 133–154. [CrossRef] [PubMed]
29. Borghesi, A.; Scrimieri, A.; Michelini, S.; Calandra, G.; Golemi, S.; Tironi, A.; Maroldi, R. Quantitative CT Analysis for Predicting the Behavior of Part-Solid Nodules with Solid Components Less than 6 mm: Size, Density and Shape Descriptors. *Appl. Sci.* **2019**, *9*, 3428. [CrossRef]
30. Alilou, M.; Beig, N.; Orooji, M.; Rajiah, P.; Velcheti, V.; Rakshit, S.; Reddy, N.; Yang, M.; Jacono, F.; Gilkeson, R.; et al. An integrated segmentation and shape based classification scheme for distinguishing adenocarcinomas from granulomas on lung CT. *Med. Phys.* **2017**, *44*. [CrossRef] [PubMed]
31. Sahu, B.; Dehuri, S.; Jagadev, A. A Study on the Relevance of Feature Selection Methods in Microarray Data. *Open Bioinf. J.* **2018**, *11*, 117–139. [CrossRef]
32. Söderman, C.; Johnsson, A.A.; Vikgren, J.; Norrlund, R.R.; Molnar, D.; Svalkvist, A.; Månsson, L.G.; Båth, M. Evaluation of Accuracy and Precision of Manual Size Measurements in Chest Tomosynthesis using Simulated Pulmonary Nodules. *Acad. Radiol.* **2015**, *22*, 496–504. [CrossRef]
33. Serra, J. *Image Analysis and Mathematical Morphology*; Academic Press: Cambridge, MA, USA, 1982.
34. Gierada, D.S.; Politte, D.G.; Zheng, J.; Schechtman, K.B.; Whiting, B.R.; Smith, K.E.; Crabtree, T.; Kreisel, D.; Krupnick, A.S.; Patterson, A.G.; et al. Quantitative Computed Tomography Classification of Lung Nodules: Initial Comparison of 2- and 3-Dimensional Analysis. *J. Comput. Assist. Tomogr.* **2016**, *40*, 589–595. [CrossRef] [PubMed]