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.

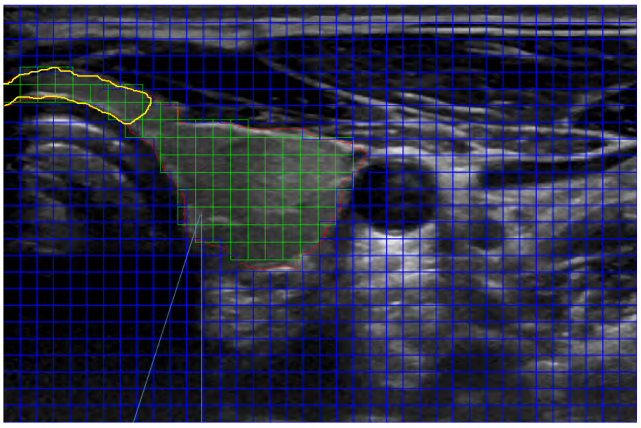

**FIGURE 1.** The figure represents the division of a 2D thyroid US images into smaller texture patches. In the figure, Green: Thyroid Patches, Blue: Non-Thyroid Patches, Red:Ground Truth and Yellow: Isthmus Region.

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).

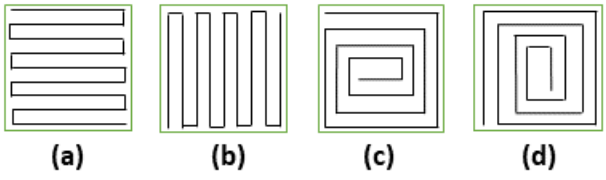


**FIGURE 2.** Conversion of texture patch to four different signals. ((a)ZigZag, (b)ZigZag 90 degree rotated, (c) Spiral and (d) Spiral 90 degree rotated respectively. Adopted from [3].

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 $\gamma$ 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 $\gamma$. 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'<br>• $C = 0.7$<br>• $\gamma = 1.0$ | • tree depth = 10<br>• minimum samples to split leaf node = 2<br>• minimum samples to split a node = 4<br>• number of trees = 200<br>• bootstrap = 'true' | • epochs = 100<br>• learning rate = 0.1<br>• momentum = 0.9<br>• 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

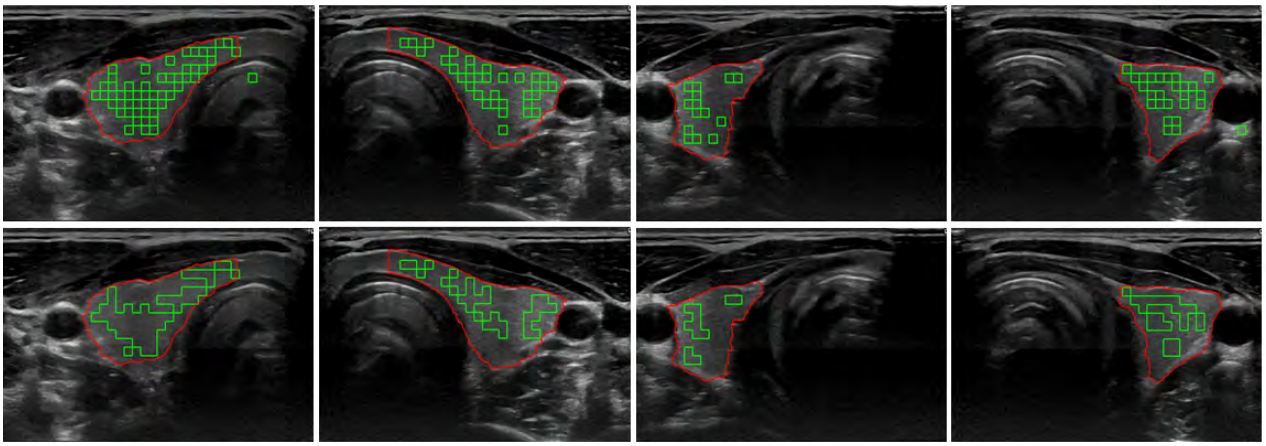

**FIGURE 4.** Examples of thyroid texture classification and segmentation using SVM and comparison with ground truth on Dataset 1.

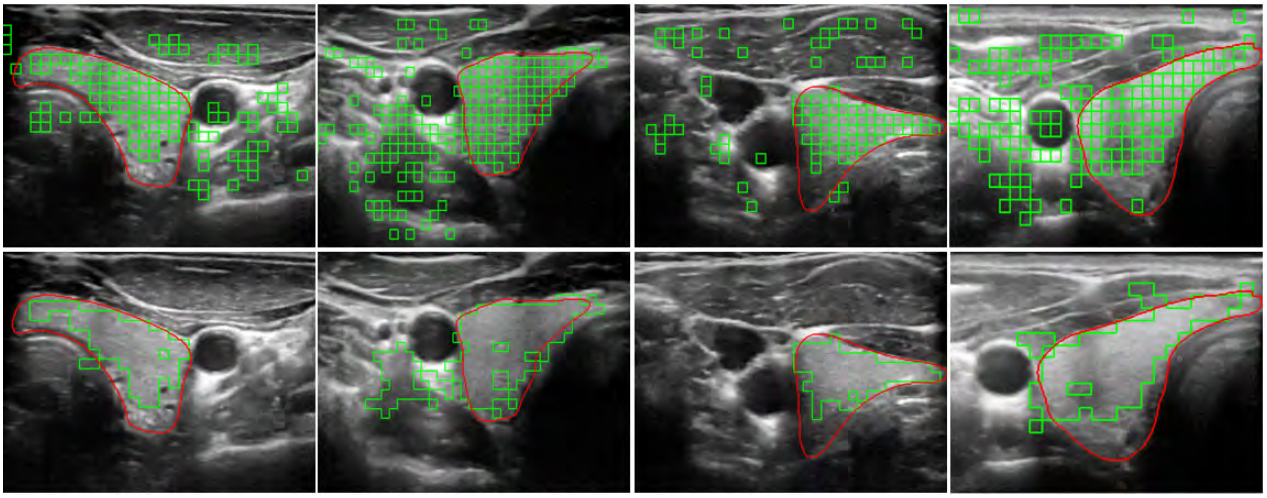

**FIGURE 5.** Examples of thyroid texture classification and segmentation using SVM and comparison with ground truth on Dataset 2.

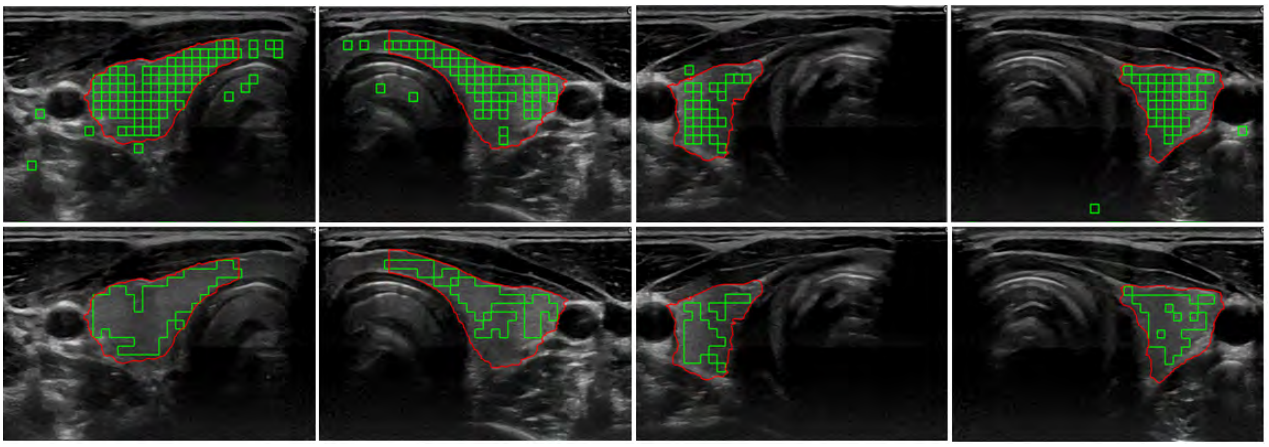

**FIGURE 6.** Examples of thyroid texture classification and segmentation using ANN and comparison with ground truth on Dataset 1.

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

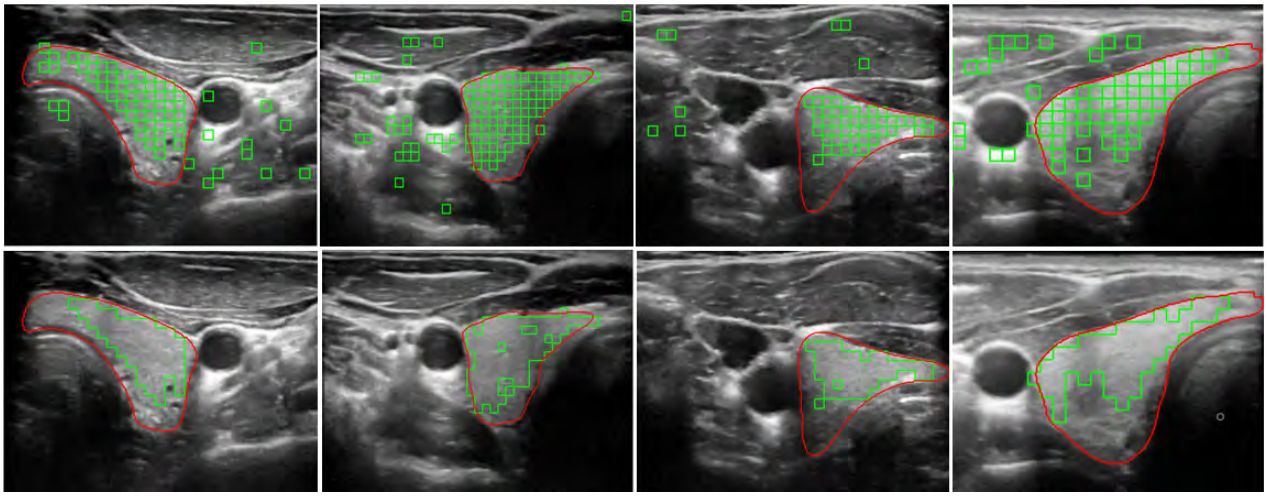

**FIGURE 7.** Examples of thyroid texture classification and segmentation using ANN and comparison with ground truth on Dataset 2.

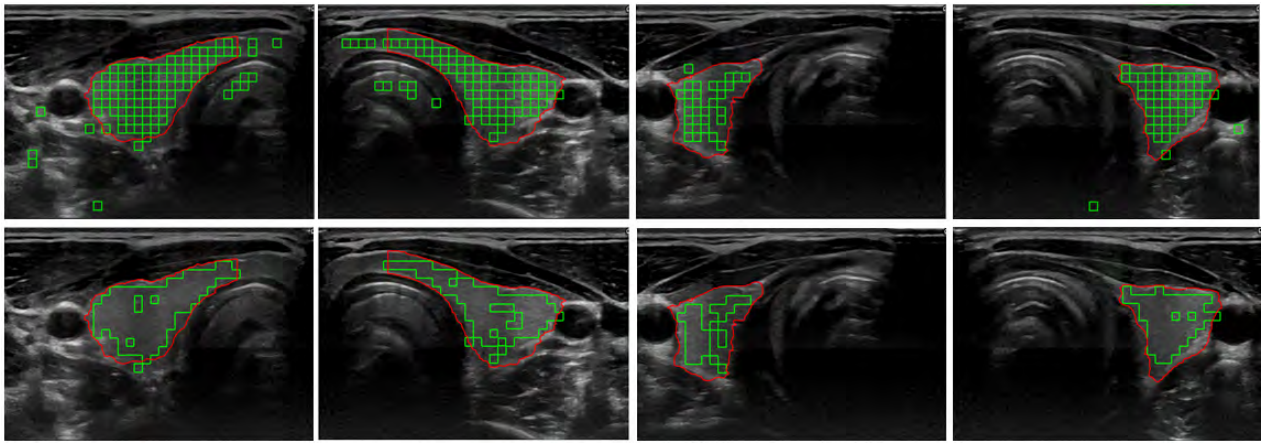

**FIGURE 8.** Examples of thyroid texture classification and segmentation using RFC and comparison with ground truth on Dataset 1.

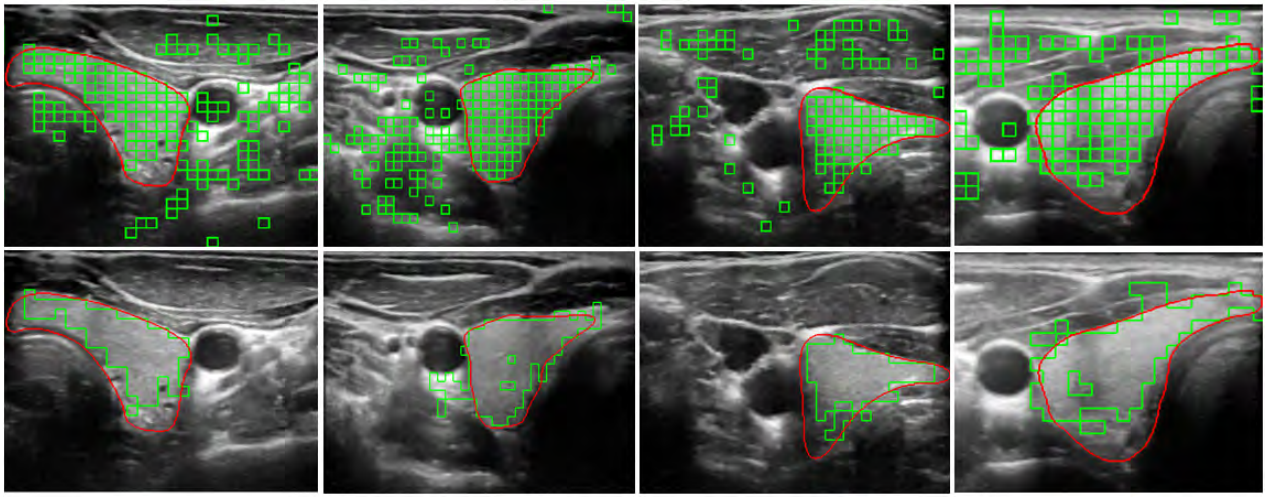

**FIGURE 9.** Examples of thyroid texture classification and segmentation using RFC and comparison with ground truth on Dataset 2.

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

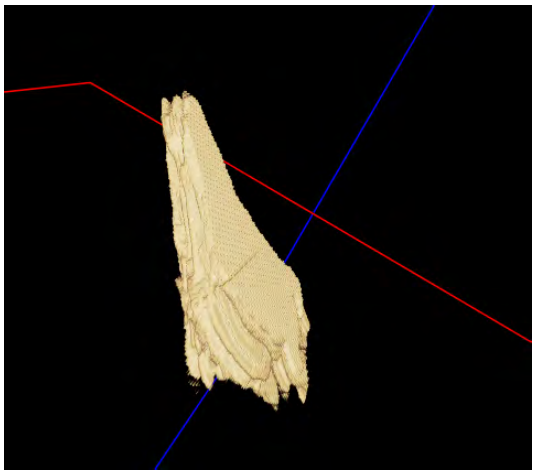

**FIGURE 10.** A 3D reconstructed thyroid after texture classification and segmentation using Imfusion.

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.