



# An Adaptive Feature Extraction Model for Classification of Thyroid Lesions in Ultrasound Images

Hatwib Mugasa<sup>a</sup>, Sumeet Dua<sup>a,\*\*</sup>, Joel E W Koh<sup>b</sup>, Yuki Hagiwara<sup>b</sup>, Oh Shu Lih<sup>b</sup>, Chakri Madla<sup>c</sup>, Pailin Kongmebhol<sup>c</sup>, Kwan Hoong Ng<sup>d</sup>, U Rajendra Acharya<sup>b,e,f</sup>

<sup>a</sup>Computer Science Program, College of Engineering and Science, Louisiana Tech University, Ruston, LA 71272 USA

<sup>b</sup>Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore 599489 (email: [aru@np.edu.sg](mailto:aru@np.edu.sg))

<sup>c</sup>Department of Radiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

<sup>d</sup>Department of Biomedical Imaging, University of Malaya, Kuala Lumpur, Malaysia

<sup>e</sup>Department of Biomedical Informatics and Medical Engineering, Asia University, Taiwan

<sup>f</sup>International Research Organization for Advanced Science and Technology (IROAST) Kumamoto University, Kumamoto, Japan

## ABSTRACT

The thyroid is the chief hormonal gland that controls the growth, metabolism, and maturation of the body. However, the function of the thyroid gland could be disrupted if it produces too much or too little hormones. Furthermore, there could be abnormal growth in thyroid cell tissue, leading to the formation of a benign or malignant thyroid lesion. Ultrasound is a typical non-invasive diagnosis approach to check for cancerous thyroid lesions. However, the visual interpretation of the ultrasound thyroid images is challenging and time-consuming. Hence, a feature engineering model is proposed to overcome these challenges. We propose to transform image pixel intensity values into high dimensional structured data set before fitting a Regression analysis framework to estimate kernel parameters for an image filter model. We then adopt a Bayesian network inference to estimate a subset for the textural features with a significant conditional dependency in the classification of thyroid lesions. The analysis of the proposed feature engineering model showed that the classification performance had an overall significant improvement over other image filter models. We achieve 96.00% classification accuracy with a sensitivity and specificity of 99.64% and 90.23% respectively for a filter size of  $13 \times 13$ . The analysis of results indicate that the diagnosis of ultrasound images thyroid nodules is significantly boosted by adaptively learning filter parameters for feature engineering model.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

The thyroid is an organ located in the lower front of the neck. It is a butterfly-shaped hormonal gland that regulates blood pressure, body temperature, heart rate, metabolism, and weight [1]. This hormone also controls the amount of calcium in the blood. It is common to have an abnormal growth of thyroid cells in the tissues of the thyroid gland, which introduces the formation of nodules in the thyroid [1]. Approximately 90% of us are likely to have at least one thyroid nodule by the time we reach 80 years old. The probability of developing a thyroid nodule regardless of be benign (non-cancerous) or malignant

(cancerous) is higher as we age [2]. It is also noted that most of the thyroid nodules are benign, while only a small percentage are malignant.

Benign thyroid nodules are non-life-threatening; hence, treatment is not necessary unless the nodule grows too big that it is difficult to breathe. Nonetheless, it is crucial to detect the malignant thyroid nodules at an early stage and to seek immediate treatment [1]. Generally, there are four main types of thyroid cancers. They are, namely, Anaplastic, Follicular, Medullary, and Papillary Thyroid cancer [1]. These four types of cancers are based on characteristics of the cancer cells when viewed under a microscope.

However, the growth of these thyroid nodules exhibits almost no symptoms [2]. Therefore, the nodules are only discovered during a routine health examination. It would probably be too

<sup>\*\*</sup>Corresponding author: Tel: +1-318-257-2871  
e-mail: [sdua@latehc.edu](mailto:sdua@latehc.edu) (Sumeet Dua)

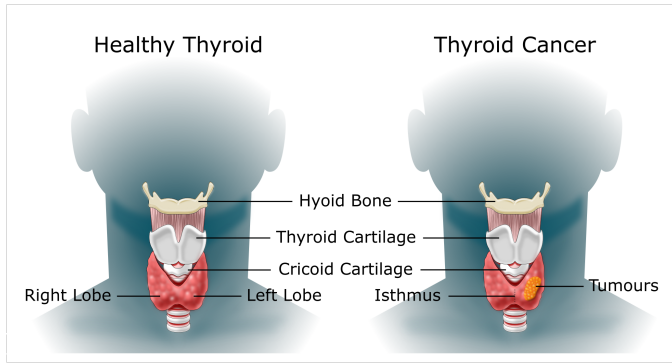


Fig. 1: An illustration of anatomical structure of healthy and cancerous thyroid.

late for treatment if the malignant thyroid nodules are detected at an advanced stage. According to the American Cancer Society, it is estimated that there are 56,870 new cases of thyroid cases in the United States presently [3]. It is reported that the number of deaths due to thyroid cancer this year is 2,010.

Fig. 1 is a representation of a healthy and cancerous thyroid. It can be seen that a cancerous thyroid has nodules growing in the thyroid gland. Whereas, there is no abnormal growth of nodules in the thyroid gland. However, not all the nodules formed in the thyroid gland is cancerous. It can be non-cancerous too. The ultrasound equipment is a standard medical imaging modality used to locate and evaluate thyroid lesions [4]. It is a safe, non-ionizing, and non-invasive technique to examine the thyroid glands [5]. However, the images (benign and malignant thyroid lesions) captured with the ultrasound machine is indistinct and undifferentiable (see Fig. 1). Thus, the visual inspection of the ultrasound images must be analyzed by ultrasonographers who have adequate experience and expertise to ensure an accurate diagnosis. Otherwise, the diagnosis of the ultrasound images may be inaccurate and subjective.

Therefore, to overcome the challenge of accurate interpretation, a computer-aided diagnosis (CAD) system is developed with a twofold objective, namely to assist ultrasonographers in giving an objective second-opinion of the thyroid lesion diagnosis (to determine if it is benign or malignant). Furthermore, to make the proposed CAD system time-efficient by minimizing any manual interpretation of the images that is laborious.

Considering the inherent high-dimension of the data, our contributions in the proposed data mining framework focuses on overcoming specific challenges. This includes the extraction of relevant texture descriptors from low-contrast, noisy images, and the reduction of inherent dependencies between features. We hypothesize that by learning the image filter parameters from a regularized linear regression model, we minimize both the variance and correlation among attributes (or features) and boost the differentiation of the different thyroid nodules (benign or malignant) through effective model building.

The goal of the proposed CAD system therefore aims at reducing the number of misdiagnoses made by ultrasonographers, and hence, minimize the need for biopsy. A biopsy is typically performed to confirm if the thyroid nodule is benign or malignant. It is an invasive approach whereby a tissue sample is taken from the thyroid gland and analyzed with a microscope

[4]. Since it is an invasive test, the preferred alternative is the non-invasive and safe approach to confirm the diagnosis made by the ultrasonographers. This, in turn, will allow hospitals to conduct a mass thyroid screening session for the public and to create awareness of thyroid cancer.

## 2. Previous Works

Feature extraction plays a vital role in the pre-processing phases of a classification problem by extracting the relevant information from the original feature. Texture feature extraction is beneficial in discovering hidden and potentially relevant patterns in images. This process consists of (i) feature construction, (ii) subset generation, (iii) feature evaluation, and (iv) assessment of the extracted features. It is also beneficial in performing dimensionality reduction and feature selection for the analysis and mining of Big data. In image processing algorithms, feature extraction is used to locate pixels on an image that are significantly similar to pixels in the images.

Deep Learning algorithms like Convolutional Neural Networks (CNN) [6] have been extremely successful in image classification applications; however, their drawback is that they work best with well-defined objects. The research motivation in this study is to optimize image filtering methods for efficient processing of low contrast images with high speckle noise levels of ultrasound thyroid nodule images.

Hailiang, et al., [7] suggested a thyroid papillary carcinoma detection method for ultrasound images by concatenating the shallow and deep layers of a CNN. Their method was inspired by a state-of-the-art Fast Region-based Convolutional Network method (Fast R-CNN) [8]. They reported an accuracy of 93.50% for detecting papillary thyroid carcinoma regions and 81.50% for detecting benign and normal tissue. A Support Vector Machines (SVM) classifier was adopted for the diagnosis of thyroid nodule [9, 10]. Vanithamani, et al., [9] adopted Fuzzy C-Means Clustering and Gray Level Co-occurrence Matrix (GLCM) framework to despeckle ultrasound images and extract image feature. Li, et al., [10] proposed a novel local feature to describe the rotation invariant properties of the texture. In their work, a pairwise rotation-invariant spatial context and image scale information are adopted to describe rotation invariant properties of the texture features. Their proposed method outperformed existing algorithms with a reported mean accuracy of 99.80%. A Principal Component Analysis (PCA) based segmentation method was proposed by Snekhalatha, et al., [11] to extract geometric and statistical features of thyroid gland ultrasound image. An interclass variance analysis method was adopted in their work to automate the segmentation of thyroid nodules. A feed forward propagation network, that achieved a sensitivity of 85.71%, specificity of 95%, and accuracy of 91%, was later applied for the classification of normal and abnormal thyroid glands.

Hybrid models like the Gabor filter method and analysis of Haralick Features [12] have recently gained popularity in diagnostic classification of images in the medical field because of the advantages of performing both structures and statistical operations. Haralick texture features have been used for the analysis of digital mammograms by comparing the performance of

simple distance-measure classifications between features [13] and to classify abnormalities in lung CT images as either having tumors or a buildup of fluid [14].

### 2.1. Image Filtering

There are various spatial and frequency domain from the output of the techniques can be performed to analyze the images. The Windowing Filter Methods [15] perform a convolution over the original image by multiplying a sliding filter of fixed size (usually square) to the image. The window filter radius  $r$  has to be uneven in order to align the filter center with the current image and ensure the convolution operation is uniformly applied to all pixels.

Given an image  $X$  of size  $N \times M$ , a Window Filter  $W$  with radius  $r$  is defines as:

$$W_r = \begin{pmatrix} w_{i-r,j-r} & \cdots & w_{i-r,j} & \cdots & w_{i-r,j+r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{i,j-r} & & w_{i,j} & & w_{i,j+r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{i+r,j-r} & \cdots & w_{i+r,j} & \cdots & w_{i+r,j+r} \end{pmatrix} \quad (1)$$

where  $3 \leq r \leq \min(N, M)$ . A convolution operation is applied to the elements  $w_{i,j}$  on a  $(2 \times r) + 1$  square region of the original image to obtain a filtered output  $x_{a,b}$ . One major limitation of performing Windowing with a filter size  $r$  is that it is operationally complex  $O(r^2)$ , and quadruples to  $O(4r^2)$  when the filter size is doubled.

Linear operations techniques like the Gaussian and Mean filter method [16] performs smoothing in windowed pixel regions to remove specific noise like Gaussian, Salt and Pepper noise. Alternatively, Nonlinear techniques like Median filter methods are used to remove impulse noise while preserving edges in the images. The complexity of the Linear technique is  $O(1)$  however, the noise is assumed to be generally distributed over zero-mean. Some of the popular variations of the Mean and Median filters are Distance Weighted Median Filter, Gaussian Filter [17], K-Nearest Neighbor Filter (KNNFilter) [18] that obtains an averages output of the k-least different pixels in the window, and the Winsorized Mean Filter [19] and Modified Trimmed Mean Filter (MTMFilter) [20] methods that both selects a subset of the windowed pixel.

Linear filters are the simplest form of filter methods because of their ease-of-use in finding areas of rapid change (edges) in images. However, they are prone to edge smoothing in a scenario where the noise is highly granular because they combine all pixels of the original image in the output value. An alternative uses complex non-linear methods that reduce noise levels without simultaneously blurring edges. Nonlinear image filtering techniques are the preferred option over linear techniques because of their ability to detect edges while reducing noise.

Convolution filters are used to perform multiplication operations on two 2-D matrices of different sizes to produce a smoothed version of the original image. The two input matrices include a kernel matrix and grayscale regions of the original image. During the convolution operation, the kernel matrix is

slide over the underlying unfiltered image to obtain corresponding output values  $O(i, j)$  such that:

$$O(i, j) = \sum_{w=1}^W \sum_{h=1}^H I(i+w-1, j+h-1)K(w, h) \quad (2)$$

where  $W$  and  $H$  are width and height of the image  $I(*, *)$ ,  $w$  and  $h$  are width and height of kernel  $K(*, *)$ , and  $i$  and  $j$  are the row and column positions of filtered output such that  $i = \{1, 2, \dots, W-w+1\}$  and  $j = \{1, 2, \dots, H-h+1\}$  Popular convolution kernels used in image pre-processing are the Gaussian kernel and the Laplacian or Laplacian of Gaussian kernel filters.

### 2.2. Gaussian Filter (Gaus)

The Gaussian kernel filters are used to eliminate noise in image processing by smoothing or blurring the image. The 2-D values in the Gaussian kernel filter are assumed to follow a Gaussian distribution with a mean centered at zero and is defined as:

$$G(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

where  $\sigma$  is the standard deviation of the distribution.

### 2.3. Laplacian of Gaussian (LoG)

Alternatively, the Laplacian kernel is used to detect edges in an image by emphasizing regions where the pixel intensity changes rapidly and is defined as:

$$L(x, y) = \frac{-1}{\pi\sigma^2} \left( 1 - \frac{x^2 + y^2}{2\sigma^2} \right) e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4)$$

where the standard deviation  $\sigma$  is also centered at zero mean.

### 2.4. Binarized Statistical Image Features (BSIF)

Binarized Statistical Image Features (BSIF) is a local image descriptor which produces binary codes for each pixel in the thyroid ultrasound image [21]. The pixels in the image are described as a binary coded string, and every individual pixel is assigned a code value where it represents the intensity distribution in the image. We then, use the image histogram to identify texture properties within the individual sub-region of the image. Given a neighborhood of  $n \times n$  pixels with a set of  $x$  linear filters of the same size, the  $x$ -bit label value of every bit in the binary code string is obtained by binarizing with a linear filter with zero threshold as seen in Equation 5 as:

$$s = Wl \quad (5)$$

where  $l$  is the  $n^2 \times n$  vector representation of the  $n \times n$  neighborhood and  $W$  is the  $x \times n^2$  matrix denoting the stack of the vector notations of the filters.

Likewise, every single bit is related with a different filter and the required length of the bit string sets the number of filters. Equation 6 shows that the  $i^{th}$  value of  $s$  is the function of the  $i^{th}$  linear filter  $w_i$  such that:

$$s_i = w_i^T l_i \quad (6)$$

The different filters are learned from a training set of image patches by maximizing the statistical independence of the filter responses. Hence, the BSIF descriptors learn and recognize the filters by Independent Component Analysis (ICA) filters. Therefore, each bit of the BSIF value can be attained through Equation 7 as:

$$B_i = \begin{cases} 1, & \text{if } s_i > 0 \\ 0, & \text{if } s_i \leq 0 \end{cases} \quad (7)$$

### 3. Methodology

In this study, we propose an adaptive convolution kernel filtering method that suppresses speckle noise and enhances edges of objects detected in the image using a deterministic approach of obtaining the kernel filter. Our method uses a trained Regression model to approximate optimal values for the kernel values. The Haralick method is applied to extract features that represent the texture characteristic of the image. We propose an image filtering method the adapts to information obtained from the image domain using a trained regression model. The proposed method summarized in Fig. 2 consists of the following steps:

- 1) Image Pre-processing
- 2) Filter Design
  - 1) Regression Model Design
  - 2) Estimate Kernel Parameters
- 3) Feature Extraction
- 4) Model Evaluation

#### 3.1. Image Pre-processing

During pre-processing, attenuation artifacts within the ultrasound images were handled applying a Contrast Limited Adaptive Histogram Equalization (CLAHE) [22]. We believe that CLAHE will enhance the contrast, thereby improving the resolution quality and reduces noise present in the image [23]. The CLAHE is a variation of the Adaptive Histogram Equalization (AHE) algorithm that performs Contrast-Limited enhancement of grid regions within an image. The intensity enhancement of each pixel in a local region is limited by the cumulative distribution function (cdf) slope within the neighborhood of regions. An interpolation of this process on the original image will result in the enhancement of each pixel relative to the noise in it's localized region, thus significantly improving the contrast and enhancing artifacts in the image. Fig. 4b, 4c and 4d are examples of the CLAHE output illustrating outlines of artifacts detected for different filter sizes from the original image in Fig. 4a.

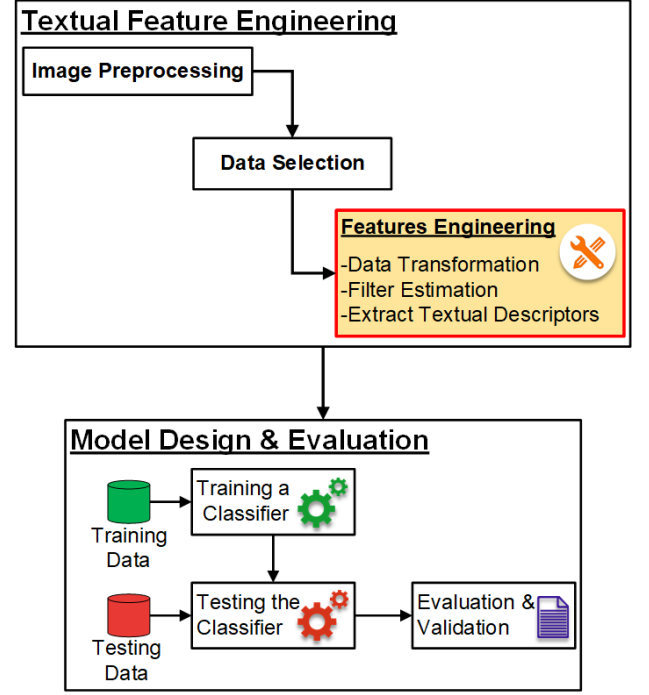


Fig. 2: The proposed block diagram to develop a CAD system for the automated distinction of benign and malignant thyroid lesions.

#### 3.2. Filter Design

The selected sampled image patch were transformed into 1-dimensional structured data set with the target response represented by the expected pixel value after applying (a) high-pass filtering within region of interest (ROI) and (b) low-pass filtering on out-of-scope regions (non-ROI) region. A linear regression model is fitted to the transformed data set to estimate model coefficient parameters. An Elastic-Net regularization path was used to control the concentration and shrinkage of coefficient parameters during the model training process.

##### 3.2.1. Regression Models

Linear Regression is used to model the relationship between independent input variables and their dependent target variables by fitting a linear equation. The association between the input and target variables can be determined by first examining the data points onto a scatterplot and analyzing their correlation coefficient. A positive correlation coefficient value indicates that there exists a positive association between the variables and negative values for the correlation coefficient indicates a negative association. The regression coefficient values can be estimated using singular values decomposition.

A Linear Regression assumes the residual error are random follows a normal distribution for input variables  $X$  and dependent variables  $Y$  such that the model equation is defined as:

$$Y_i = \beta_0 + \beta X_{i,*} + \epsilon_i \forall j \in 1, \dots, p \quad (8)$$

$$= \beta_0 + \beta_1 X_{i,1} + \dots + \beta_p X_{i,p} + \epsilon_i$$

where  $i$  are the number of observations for  $i = 1, \dots, n$ ,  $j$  are the features for  $j = 1, \dots, p$ ,  $\beta = (\beta_1, \dots, \beta_p)$  are coefficient

parameter  $\beta$  and  $\epsilon_i \sim N(0, \sigma^2)$  are the residual error parameters with a probability density function:

$$f(y_i) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(y_i - X_{i,*})^2}{2\sigma^2}\right] \quad (9)$$

The prediction of  $Y_i$  of the regression model is equal to:

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y \quad (10)$$

#### Singular Value Decomposition (SVD)

Singular Value Decomposition proposed in [24] is a matrix decomposition method commonly used in dimensionality reduction in high dimensional and strongly collinearity matrix. The singular value decomposition of a matrix  $A$  is defined as:

$$A = UDV^T \quad (11)$$

where  $D$  an  $n \times n$  diagonal matrix with the singular values,  $U$  an  $n \times n$  matrix of left singular vectors,  $V$  a  $p \times n$  matrix of right singular vectors, and columns of  $U$  and  $V$  are orthonormal, i.e:

$$UU^T = I = VV^T \quad (12)$$

The Equation 10 can be rewritten in terms of the SVD-matrices in Equation 11 to estimate coefficient parameters  $\hat{\beta}$  of the linear regression model as:

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T Y \\ &= (X^T X + \lambda I)^{-1} X^T Y \end{aligned} \quad (13)$$

where  $\lambda$  is the tuning parameter (Ridge penalty parameter),  $\lambda \in [0, \infty)$  and  $I$  is a  $(p \times p)$ -dimensional identity matrix. The Newton-Raphson equation is applied iteratively to find the roots of the log-likelihood estimation function in Equation 8 to derive new values of the coefficient as:

$$\begin{aligned} \hat{\beta}_{new} &= \hat{\beta}_{old} + (X^T W X)^{-1} X^T [Y - g^{-1}(X; \beta)] \\ &= (X^T W X)^{-1} X^T W \{X\hat{\beta}_{old} + W^{-1} [Y - g^{-1}(X; \beta)]\} \\ &= (X^T W X)^{-1} X^T W Z \\ &= \arg \max_{\beta} (Z - X\beta)^T X^T W (Z - X\beta) \end{aligned} \quad (14)$$

where  $W$  is a  $(n \times n)$ -dimensional diagonal matrix with  $W_{i,i} \in [0, 1]$  representing the weight of the  $i^{th}$  observation,  $Z$  is the adjusted response  $Z = X\hat{\beta}_{old} + W^{-1} [Y - g^{-1}(X_{i,*}; \beta^{old})]$ . The Link function  $g(\cdot)$  is a function that links responses to observed variables  $X_{i,*}$  such that:

$$p(X_{i,*}) = p(Y_i) = g(X_{i,*}; \beta)^{-1} = \frac{\exp(X_{i,*}; \beta)}{1 + \exp(X_{i,*}; \beta)} \quad (15)$$

The likelihood estimates of the Linear Regression is thus defined as:

$$L(Y|X; \beta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(y_i - X_{i,*}\beta)^2}{2\sigma^2}\right] \quad (16)$$

and is rescaled to a log-likelihood estimation as:

$$\mathcal{L}(Y|X; \beta) = -n \log \sqrt{2\pi\sigma} - \frac{1}{2\pi\sigma^2} \sum_{i=1}^n (y_i - X_{i,*}\beta)^2 \quad (17)$$

#### Logistic Regression

Logistic Regression or Logit Regression is a type of probabilistic statistical classification model used to explain the response of binary targets. The Logistic Regression algorithm iterative applies a Maximum Log-likelihood estimator to obtain a set combination of the coefficients of the observations. The coefficients are the fit to a regression loss function that approximates the probability of the discrete target value. Because Linear Regression functions are unbound, a logarithmic transformation is performed to obtain a Logistic Regression function of the form:

$$\begin{aligned} Y_i &= \beta_0 + X_{i,*}\beta_i \rightarrow \ln \left[ \frac{p(X_{i,*})}{1 - p(X_{i,*})} \right] \\ &= \beta_0 + X_{i,*}\beta_i, \forall p(X_{i,*}) \in [0, 1] \end{aligned} \quad (18)$$

where  $p(X_{i,*})$  is logistic link function that links the response values to input variables.

A numerical root-finding method, like the Newton-Raphson method, is used interactively to approximate the coefficient parameters  $\beta$  by converging to a maximum log-likelihood estimate. The Likelihood estimation function of the Logistic Regression model is defined as:

$$L(Y|X; \beta) = - \prod_{i=1}^n [(p(y = 1|X_{i,*}))^{Y_i} (p(y = 0|X_{i,*}))^{1-Y_i}] \quad (19)$$

Equation 19 is rescaled using logarithm to reduce the likelihood estimations being skewed towards large values. This is known as the log-likelihood estimation function as  $\mathcal{L}(Y|X; \beta)$  denoted as:

$$\mathcal{L}(Y|X; \beta) = -\frac{1}{n} \sum_{i=1}^n [Y_i \cdot (\beta_0 + X_{i,*}^T \beta)] \quad (20)$$

$$+ \ln(1 + \exp(\beta_0 + X_{i,*}^T \beta))] \quad (21)$$

#### Regularization

The model coefficients  $\hat{\beta}$  of high dimensional data are known to be highly sparse and tend to have an infinite number of possible estimations for a single coefficient. Regularization controls the coefficients by modeling an objective path that reduces both the variance and the Sum of Square Errors (SSE) between the observed and predicted targets. This is achieved by applying a penalty factor to the log-likelihood function of a Logistic Regression model as denoted below:

$$\mathcal{L}_*(Y|X; \beta) = \max_{\beta_0, \beta} \{\mathcal{L}(Y|X; \beta) + Obj(P; \beta)\} \quad (22)$$

where  $Obj(P; \beta)$  is the objective function for setting a penalty parameters  $P$ .

#### Ridge Regularization

The Regularization paths proposed to handle such challenges include the Ridge, Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net Regularization Paths. The



Ridge Regularization Path shrinks correlated coefficients towards a common estimated value and its objective function is denoted as:

$$\mathcal{L}_{Ridge}(Y|X;\beta) = \mathcal{L}(Y|X;\beta) + \lambda \cdot \frac{1}{2} \sum_{j=1}^p \beta_j^2 \quad (23)$$

*Least Absolute Shrinkage and Selection Operator Regularization(LASSO)*

The Least Absolute Shrinkage and Selection Operator (LASSO) Regularization path eliminates irrelevant coefficients from the model its objective function is defined as:

$$\mathcal{L}_{LASSO}(Y|X;\beta) = \mathcal{L}(Y|X;\beta) + \lambda \cdot \sum_{j=1}^p |\beta_j| \quad (24)$$

where  $\lambda \in \mathbb{R}^+$  set to control the  $\sum_{j=1}^p |\beta_j|$  (LASSO) L1 normalization and  $\sum_{j=1}^p \beta_j^2$  (Ridge) L2 normalization.

*Elastic Net Regularization*

The Elastic Net Regularization sets a balance between the Ridge and LASSO paths by introducing of a trade-off parameter  $\alpha, 0 \leq \alpha \leq 1$ . The path taken by the Elastic Net will be similar to that of a Ridge Path if  $\alpha$  is equal to 1 and a path similar to LASSO if  $\alpha$  is equal to 0. The log-likelihood function for the Elastic Net Regularization is denoted as:

$$\begin{aligned} \mathcal{L}_{ENet}(Y|X;\beta) &= \mathcal{L}(Y|X;\beta) \\ &+ \lambda \cdot [(1 - \alpha) \frac{1}{2} \sum_{j=1}^p \beta_j^2 + \alpha \cdot \sum_{j=1}^p |\beta_j|] \end{aligned} \quad (25)$$

*Coefficient Shrink Factor*

In addition to the Ridge and LASSO factors, a rate of shrink factor of each coefficient can be set to the objective function  $Obj(P_{\lambda,\alpha};\beta)$  to control the penalty ratio of each coefficient individually. The Objective function in an Elastic Net Regularization model can be transformed into a weighted Objective function as:

$$Obj(P_{\lambda,\alpha};\beta) = \lambda \sum_{j=1}^p \omega_j [(1 - \alpha) \frac{1}{2} \beta_j^2 + \alpha |\beta_j|] \quad (26)$$

where where  $\omega_j$  is the rate of shrink of coefficient  $j$ , for  $0 \leq \omega_j \leq 1$

### 3.2.2. Estimating Kernel Parameters

Variations of the filter kernel are generated from the estimated model coefficients that were achieved from adjusting the  $L_1$  and  $L_2$  penalties by applying Elastic Net regularization and minimizing the root square error. A set of Kernels were obtained from sampling different image patches sizes (i.e., 5, 7, ..., 29) corresponding to the sizes of the resultant kernels. The resulting filters were applied to the images to extract Haralick texture features illustrated in Fig. 4.

### 3.3. Feature extraction

The Haralick texture features method was used in this work to identify and extract the texture features of the ultrasound

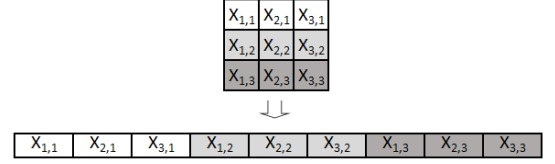


Fig. 3: Transformation of an image patches to 1-D data frame.

image. Haralick texture features are statistical calculations of an image or region of interest (ROI) in the image based on spatial dependencies of a Gray Level Co-occurrence Matrix (GLCM) using four angular mean. This approach assumes that spatial dependence frequencies between neighboring cells in a grayscale image are a function of the angular relationship and their distance. Geometrical relationships of GLCM measurements are applied to four-radian angles  $\theta = 0, \pi/4, \pi/2, \pi/4$  and the distance  $d = \max\{|x|, |y|\}$  to adjacent neighboring cells.

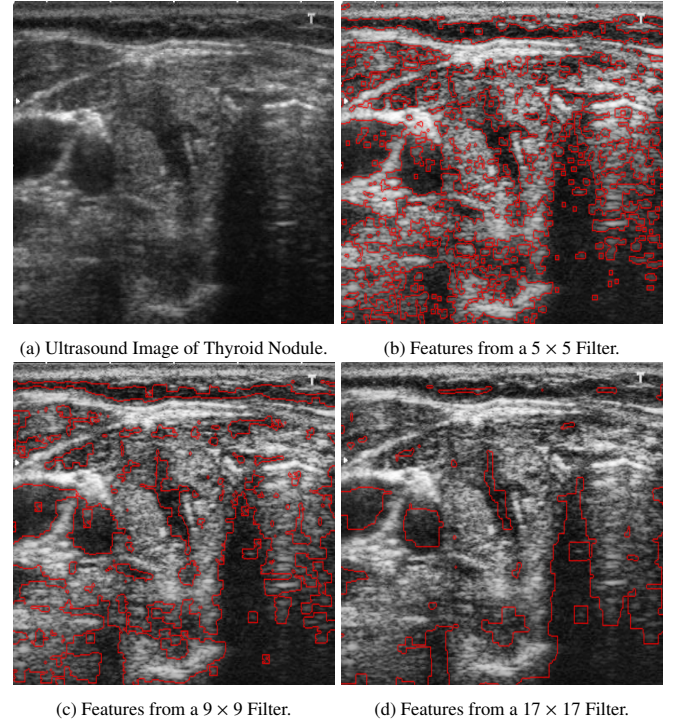


Fig. 4: Outlines of Haralick features extracted from applying (4b)  $5 \times 5$  Filter, (4c)  $9 \times 9$  Filter, and (4d)  $17 \times 17$  Filter.

## 4. Feature selection

We performed partial dependence test of each selected Haralick features estimated from the single class and multilabel class models Fig. 5 below. The partial dependence is the approximation of the relationship between a subset of the feature space. Feature importance measures were also computed by applying the Boruta algorithm with the Random Forest (RF) classifier of 500 trees and  $\sqrt{n}$  randomly selected features for each tree. On completion, the Boruta algorithm categorized the features



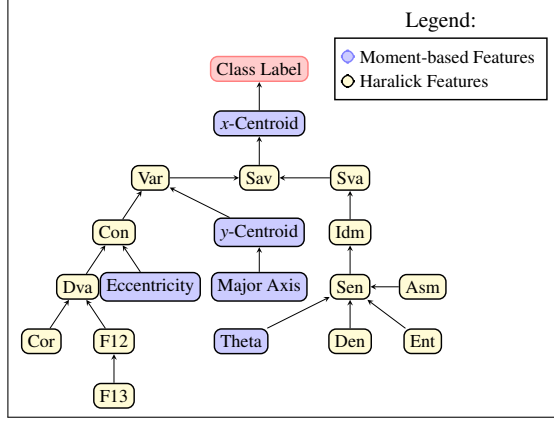


Fig. 9: Bayesian network of Moment feature properties.

S.

$$I_G(S, A) = I_E(S) - \sum_{i=1}^n \left[ p(\subset_S^{A_i}) \times I_E(\subset_S^{A_i}) \right] \quad (27)$$

where;

- $I_E(S)$  and  $I_E(\subset_S^{A_i})$  is the Shannon Entropy of set  $S$  and a subset  $A_i$  of  $S$  respectively, and
- $p(\subset_S^{A_i})$  is the probability of the instance in a subset  $A_i$  of  $S \rightarrow \frac{\text{size}(\subset_S^{A_i})}{\text{size}(S)}$ .

## 5. Experiment

In the study, images of benign and malignant thyroid lesions were obtained from the Computer Imaging and Medical Applications Laboratory (CIM@LAB) systems which is an open access dataset at the National University of Colombia - Universidad Nacional de Colombia (UNC) and Instituto de Diagnostico Medico (IDIME) [25]. A total of 388 individual patient cases with 400 diagnostic observations were used in this study. Each observation was saved in an uncompressed JPEG image. Patient cases were annotated by expert radiologists, and the information was saved in an XML file. This information included age, gender, description of the thyroid nodule, shape, margin, calcification, echogenicity, and a Thyroid Imaging Reporting and Data System (TI-RADS) scale described in Table 2. The TI-RADS lexicons 1, 2 & 3 were categorized as benign while 4a, 4b, 4c & 5 were categorized as malignant.

New cases with respective patient information are uploaded to the database periodically. A typical benign and malignant thyroid lesion image is represented in Fig. 10. These images were captured with various ultrasound equipment namely the TOSHIBA Nemio 30 and a TOSHIBA Nemio MX Ultrasound linear and convex transducers set at 12MHz. Hence, not at all the ultrasound scans from every subject uses the same ultrasound machine.

We assumed the data followed a Gaussian distribution and scaled it to a standard normal distribution with a mean of 0 and a standard deviation of 1. We partitioned the data into a 60/40 split for training set and testing sets. Random samples

Table 2: The Thyroid Imaging Reporting and Data System (TI-RADS) scale used to describe Thyroid lesions.

Scale	Description	Classification
1	Normal thyroid	Benign
2	Benign Follicular Nodule	Benign
3	No suspicious lesion	Benign
4a	1 suspicious lesion	Malignant
4b	2 suspicious lesion	Malignant
4c	3 or 4 suspicious lesions	Malignant
5	5 or more lesions	Malignant

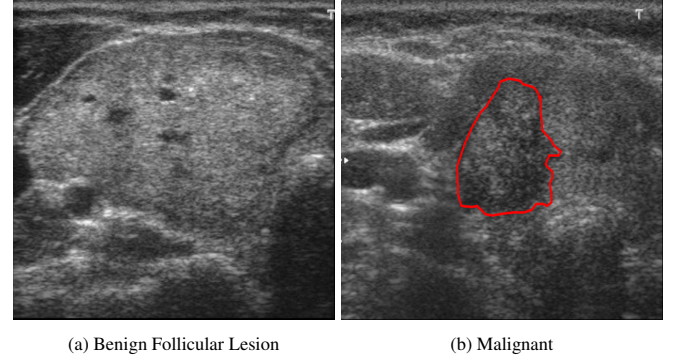


Fig. 10: Ultrasound images of (a) Benign and (b) Malignant Thyroid Lesions with the Regions of Interest (ROI) in red.

non-overlapping image patch from a training set were then selected to be fitted to a Linear Regression model. The training set was encapsulated inside a 10-fold cross validation where 9 folds were fitted to the model and one set retained for validation of the model. The cross-validation process was then repeated 10 times. Estimated model coefficients at varying  $L_1$  and  $L_2$  configurations of an Elastic Net regularization were structured into a matrix of filter kernels.

## 6. Model Validation

We applied the filtering of the data with filters kernels constructed from the proposed method and calculated gray-level co-occurrence matrices using the Haralick texture features. A subset of features from images with a significant height-to-width ratio were then filtered and fitted to a classifier. A comparison of the proposed model with the Binarized Statistical Image Features (BSIF), Gaussian and Laplacian of Gaussian extraction methods was performed using a Random Forest and Recursive Partitioning classifiers on the training set.

The 40 split disjointed testing data set was then used to evaluate the proposed model and a classification performance was used to evaluate it. We fitted the independent testing data set to Recursive Partitioning and Random Forest Classifiers and evaluated their performance by measuring the Classification Accuracy and Specificity (1-False-Positive Rate) against the Sensitivity (True-Positive Rate). Thus, Specificity is 1 - (minus) the proportion of the number of benign samples that were identified



Table 3: Evaluation of the proposed filter method fitted to various classifiers.

Classifiers:	Feature Extraction Method:					
	Gaus			LoG		
	Haralick	Shape	Moment	Haralick	Shape	Moment
Random Forest	0.79	0.69	0.70	0.55	0.64	0.69
Recursive Partitioning	0.84	0.80	0.75	0.80	0.59	0.50
Naive Bayes	0.87	0.79	0.87	0.58	0.75	0.77
Logistic (GLM)	0.75	0.69	0.87	0.77	0.65	0.72
Support Vector Machine	0.48	0.47	0.42	0.57	0.61	0.68
k-NN (k = 10)	0.59	0.69	0.60	0.70	0.82	0.70

Table 4: Evaluation of the proposed filter method fitted to various classifiers (continued).

Classifiers:	Feature Extraction Method:					
	BSIF			Proposed Method		
	Haralick	Shape	Moment	Haralick	Shape	Moment
Random Forest	0.87	*0.90	0.81	0.88	0.67	0.70
Recursive Partitioning	0.84	0.84	0.87	0.84	0.75	0.41
Naive Bayes	0.89	0.78	0.67	0.97	0.71	0.77
Logistic (GLM)	0.65	0.84	0.74	0.87	0.81	0.78
Support Vector Machine	0.67	0.42	0.47	0.68	0.72	0.58
k-NN (k = 10)	0.68	0.75	0.64	*0.96	0.68	0.86

as malignant, while the Sensitivity is the proportion of malignant samples that are correctly identified as malignant.

## 7. Results

The analysis of the results obtained below show the classification performance from BSIF, Gaussian, and Laplacian filters against the proposed Adaptive Filter method in Fig. 11 and Fig. 12. Before performing feature selection, we obtained an overall accuracy of 75% with the Random Forest classifier and 83.33% with k-NN classifier where k=10 as illustrated in Table 3 and Table 4. After feature selection, the overall Accuracy of 92.88% with Random Forest and 86.21% with Recursive Partitioning were obtained. The Area Under the Curve in Fig. 13 was used to plot the Specificity and Sensitivity rates for the Random Forest classifier model for different filter sizes. A threshold is a percentage cut-off point that is used to make a decision when to classify a sample as positive (malignant). A low threshold value would increase the probability of a positive class detection (high sensitivity) but it would also increase the false alarms (low specificity). Our model illustrates that a decision threshold of 0.99 will archive a 99.64% Sensitivity and 90.23% Specificity measure.

We obtained the AUC to quantify the wellness of our model in distinguishing the two predictive values. The Elastic-Net parameters  $\alpha = 0, 0.1, \dots, 0.1$  and  $\lambda = 0, 0.5 \times 10^2, \dots, 10^3$  were used to generate kernel values for filter sizes =  $(5 \times 5), (7 \times 7), \dots, (17 \times 17)$ . The AUC of 99.00% for the  $(13 \times 13)$  Filter represents the probability for the model to correctly separate the benign cancer samples from malignant cancer.

We also used the AUC to quantify the wellness of our model in distinguishing the two predictive values. The AUC of 99.00% for the  $13 \times 13$  Filter represents the probability for the model to correctly separate the benign cancer samples from malignant cancer.

An analysis of these results show comparable improvement in accuracy, sensitivity and specificity performance of the classifiers relative with the increase in filter kernel size of the image filter for the proposed feature engineering model. However,

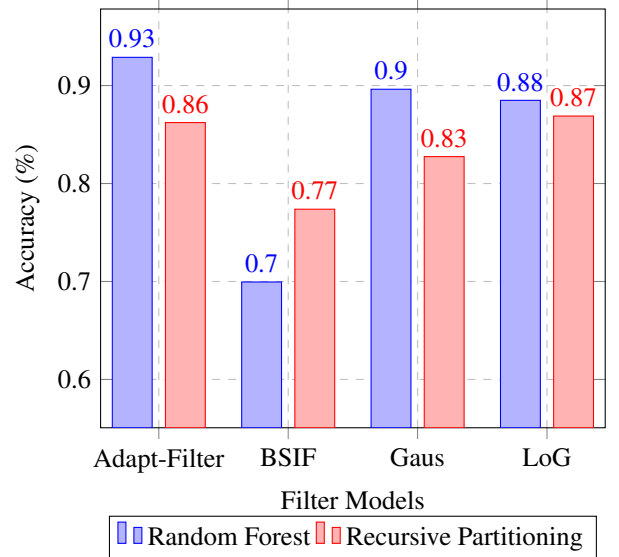


Fig. 11: Filter performances with the Random Forest and Recursive Partitioning classifiers.

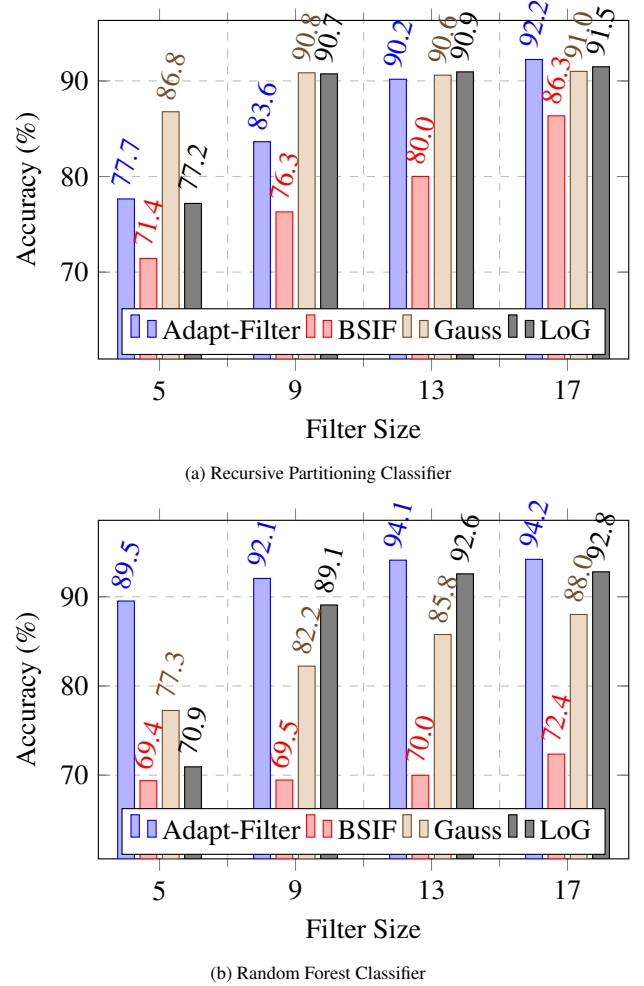
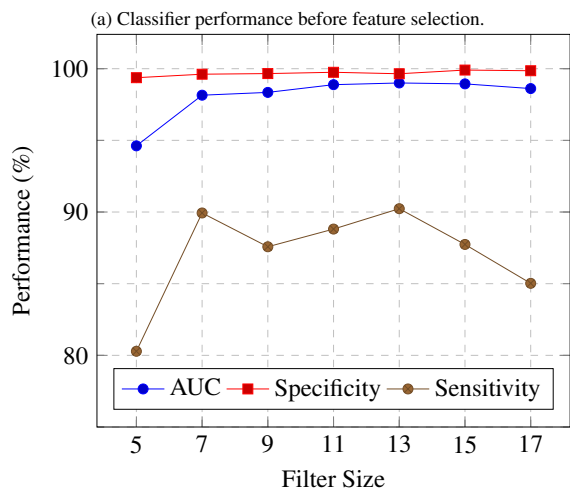
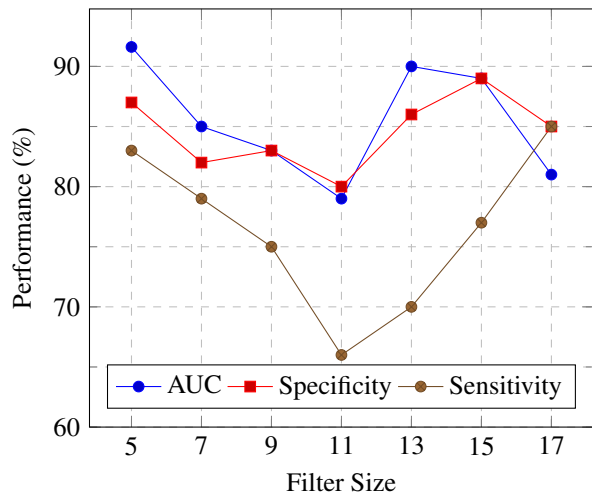


Fig. 12: Model Performance of (12a) Recursive Partitioning and (12b) Random Forest Classifiers using Image Filters from size 5 to 17.



(b) Classifier Performance after feature selection..

Fig. 13: Random Forest Classifier Performances at multiple filter sizes.

there is no significant performance gain for filter sizes greater than  $15 \times 15$  filter. We observed a gradual decline in classifier performance and a significant increase in the computational time for filter sizes higher than  $17 \times 17$ .

Table 5 shows a summary of different CAD algorithms which researchers have employed in the study of distinguishing the benign from malignant thyroid lesions with ultrasound images. Most of the researchers extracted highly significant textural features [26, 27, 28, 29, 30] for the classification of the two classes (benign and malignant). Statistical [29] and higher-order statistical [31] based features are also extracted from the thyroid images. Discrete wavelet transform (DWT) is also a technique employed by researchers to denoise the thyroid images before the extraction of features [32, 27].

## 8. Discussion

Further, an integrated index namely the thyroid malignancy index with 3D High Resolution Ultrasound (HRUS) & Contrast-Enhanced Ultrasound (CEUS) image data sets [32, 27, 28] and the thyroid clinical risk index [30] is formulated for a quick differentiation of benign or malignant thyroid lesions

Table 5: Selected studies on the CAD system for automated diagnosis of thyroid lesions (benign and malignant) with ultrasound images.

Authors (Year)	Number of images	Techniques	Performance
Ding et al. (2011) [29]	Benign: 69 Malignant: 56	Statistical and textural features, support vector machine classifier	Acc = 93.60% Sen = 94.60% Spec = 92.80%
Acharya et al. (2011) [32]	Benign: 400 Malignant: 400	DWT, KNN classifier, thyroid malignancy index	Acc = 98.90% Sen = 98.00% Spec = 99.80%
Acharya et al. (2012a) [28]	<b>HRUS</b> Benign: 400 Malignant: 400 <b>CEUS</b> Benign: 400 Malignant: 400	Grayscale texture thyroid malignancy index with HRUS and CEUS data sets	HRUS Acc = 100.00% CEUS Acc = 98.10%
Acharya et al. (2012b) [27]	Benign: 400 Malignant: 400	DWT, texture features, perceptron classifier, thyroid malignancy index	Acc = 100.00% Sen = 100.00% Spec = 100.00%
Acharya et al. (2012c) [31]	Benign: 400 Malignant: 400	Higher-order statistics based, fuzzy classifier	Acc = 99.10% Sen = 99.80% Spec = 98.50%
Acharya et al. (2016) [26]	Benign: 211 Malignant: 31	Gabor transform, textural features, decision tree classifier	Acc = 94.30%
Raghavendra et al. (2017) [30]	Benign: 211 Malignant: 31	Spatial gray level dependence and fractal texture features, thyroid clinical risk index	Acc = 97.52% Sen = 90.32% Spec = 98.57%
Chi et al. (2017) [6]	<b>Public:</b> Benign: 71 Malignant: 357 <b>Private:</b> Benign: 129 Malignant: 35	Convolutional neural network, random forest classifier	<b>Public:</b> Acc = 98.29% Sen = 99.10% Spec = 93.90% <b>Private:</b> Acc = 96.34% Sen = 86.00% Spec = 99.00%
Present work	Benign: 246 Malignant: 142	Adaptive Filter Method with Haralick Texture Features, Random Forest and Recursive Partitioning Classifiers	Acc = 96.00% Sen = 99.64% Spec = 90.23%

using a range of numerical values. Also, lately, a convolutional neural network (CNN) model is implemented to distinguish benign and malignant thyroid lesions [6]. CNN is one of the promising machine learning techniques which requires the least pre-processing with no features extraction or selection and classifier [33], [34], [35]. This model can self-learn the necessary representations for the classification of the two classes (benign and malignant) through the different layers in the CNN architecture [33]. However, Chi et al. [6] incorporated the RF classifier to classify thyroid lesions. They did not fully utilize the benefits of a CNN model.

Unlike traditional filtering methods that use heuristic methods of determining the filter parameters, we propose an adaptive method by fitting a Linear Regression Model on a subset of image pixel data and constructing the filter kernel from the coefficients estimated from the fitted model.

From the best of the authors' knowledge, no study has been done on the differentiation of benign and malignant thyroid lesions with CNN. Nevertheless, recently Ma et al. [36] adopted the CNN model to automatically identify thyroid nodules in ultrasound images. They achieved a high detection performance of 98.51%. However, their proposed CNN architecture could only pick up thyroid nodules from normal thyroid nodules. They did not extend to further categorize their thyroid nodules into either benign or malignant class. Therefore, the authors in-

tend to adopt the CNN model to categorize the thyroid lesions into benign or malignant class in future studies [34, 37]. CNN model is not employed in this work mainly due to the lack of images for both benign and malignant thyroid lesions. Big data is required to yield maximum performance with CNN model.

## 9. Conclusion

In this work, a novel feature engineering model for extracting texture descriptors to differentiate the benign and malignant thyroid nodule. A high performance of 96.00% accuracy, with a sensitivity and specificity of 99.64% and 90.23% respectively is reported for Filter size  $13 \times 13$  in the differentiation of benign and malignant thyroid using the proposed framework. Nevertheless, the proposed methodology can be further improved for maximum detection accuracy and then be considered for implementation in the hospitals. This CAD system can be easily installed in the hospitals as it is portable and cost-effective. Also, hospitals can organize a mass screening for thyroid nodules annually for an early detection so that the thyroid nodules can be treated at an earlier stage. This can help to decrease the number of end-stage thyroid cancers in the society.

## Acknowledgments

Authors thank Chiang Mai University Hospital, Thailand for providing the clinical ultrasound thyroid images.

## References

- [1] N. I. of Health, Thyroid cancer-patient version (2017). URL <https://www.cancer.gov/types/thyroid>
- [2] J. Norman, G. Clayman, Thyroid nodules: Hyperthyroidism and thyroid cancer (2017). URL <https://www.endocrineweb.com/conditions/thyroid/thyroid-nodules>
- [3] R. L. Siegel, K. D. Miller, A. Jemal, Cancer statistics, 2017, CA: A Cancer Journal for Clinicians 67 (1) (2017) 7–30. arXiv:arXiv:1011.1669v3, doi:10.3322/caac.21387. URL <http://doi.wiley.com/10.3322/caac.21387>
- [4] A. T. Association, Thyroid nodules (2017). URL [https://www.thyroid.org/wp-content/uploads/patients/brochures/Nodules\\_{ }brochure.pdf](https://www.thyroid.org/wp-content/uploads/patients/brochures/Nodules_{ }brochure.pdf)
- [5] V. Chan, A. Perlas, Basics of ultrasound imaging, in: Atlas of Ultrasound Guided Procedures in Interventional Pain Management, Springer, 2011, pp. 13–19. arXiv:arXiv:1011.1669v3, doi:10.1007/978-1-4419-1681-5\_2.
- [6] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, M. Eramian, Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network, Journal of Digital Imaging 30 (4) (2017) 477–486. doi:10.1007/s10278-017-9997-y. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5537102/>
- [7] H. Li, J. Weng, Y. Shi, W. Gu, Y. Mao, Y. Wang, W. Liu, J. Zhang, An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images, Scientific Reports 8 (1) (2018) 1–12. doi:10.1038/s41598-018-25005-7. URL <http://dx.doi.org/10.1038/s41598-018-25005-7>
- [8] R. Girshick, Fast R-CNN (Sep. 2015). arXiv:1504.08083v2. URL <http://arxiv.org/abs/1504.08083v2>; <http://arxiv.org/pdf/1504.08083v2>
- [9] R. Vanithamani, R. Dhivya, Thyroid Nodule Classification in Medical Ultrasound Images, Springer, 2018. doi:10.1007/978-3-319-60618-7\_50. URL [http://dx.doi.org/10.1007/978-3-319-60618-7\\_50](http://dx.doi.org/10.1007/978-3-319-60618-7_50)
- [10] L. Bi, Z. Shuang, Diagnosis of Thyroid Nodules Based on Local Non-quantitative Multi-Directional Texture Descriptor with Rotation Invariant Characteristics for Ultrasound Image, Journal of Medical Systems 43 (7) (2019) 1–10. doi:10.1007/s10916-019-1373-7. URL <http://dx.doi.org/10.1007/s10916-019-1373-7>
- [11] U. Snekhalatha, V. Gomathy, Ultrasound Thyroid Image Segmentation, Feature Extraction, and Classification of Disease Using Feed Forward Back Propagation Network, Springer, 2018. doi:10.1007/978-981-10-6872-0\_9. URL [http://dx.doi.org/10.1007/978-981-10-6872-0\\_9](http://dx.doi.org/10.1007/978-981-10-6872-0_9)
- [12] R. M. Haralick, K. Shanmugam, I. Dinstein, Textural Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6) (1973) 610–621. doi:10.1109/TSMC.1973.4309314.
- [13] K. Bhamidipati, V. arjunan, R. Saha, S. Selvan, Using Haralick Features for the Distance Measure Classification of Digital Mammograms, International Journal of Computer Applications.
- [14] N. Zayed, H. A. Elnemr, Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities (2015). doi:10.1155/2015/267807. URL <https://www.hindawi.com/journals/ijbi/2015/267807/cta/>
- [15] J. I. Jackson, C. H. Meyer, D. G. Nishimura, A. Macovski, Selection of a Convolution Function for Fourier Inversion Using Gridding, IEEE Transactions on Medical Imaging doi:10.1109/42.97598.
- [16] T. P. Weldon, W. E. Higgins, D. F. Dunn, Efficient Gabor filter design for texture segmentation, Pattern Recognition doi:10.1016/S0031-3203(96)00047-7.
- [17] H. Zhang, S. Krooswyk, J. Ou, Channel modeling and simulation, in: High Speed Digital Design, Elsevier, 2015, pp. 117–162. doi:10.1016/B978-0-12-418663-7.00003-4. URL <https://linkinghub.elsevier.com/retrieve/pii/B9780124186637000034>
- [18] F. Gallegos-Funes, V. Ponomaryov, S. Sadovnychiy, L. Nino-de Rivera, Median M-Type K-Nearest Neighbour (MM-KNN) Filter to Remove Impulse Noise from Corrupted Images, Electronics Letters 38 (2002) 786–787. doi:10.1049/el:20020567. URL [https://www.researchgate.net/publication/3384728\\_Median\\_M-type\\_K-nearest\\_neighbour\\_MM-KNN\\_filter\\_to\\_remove\\_impulse\\_noise\\_from\\_corrupted\\_images](https://www.researchgate.net/publication/3384728_Median_M-type_K-nearest_neighbour_MM-KNN_filter_to_remove_impulse_noise_from_corrupted_images)
- [19] K. Vasanth, T. G. Manjunath, S. N. Raj, A Decision Based Unsymmetrical Trimmed Modified Winsorized Mean Filter for the Removal of High Density Salt and Pepper Noise in Images and Videos, Procedia Computer Science 54 (2015) 595–604. doi:10.1016/j.procs.2015.06.069. URL <http://www.sciencedirect.com/science/article/pii/S1877050915013939>
- [20] Y. Yue, M. M. Croitoru, A. Bidani, J. B. Zwischenberger, J. W. Clark, Nonlinear multiscale wavelet diffusion for speckle suppression and edge enhancement in ultrasound images, IEEE transactions on medical imaging 25 (3) (2006) 297–311. doi:10.1109/TMI.2005.862737.
- [21] J. Kannala, E. Rahtu, BSIF: Binarized statistical image features, in: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), 2012, pp. 1363–1366.
- [22] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations., Computer vision, graphics, and image processing 39 (3) (1987) 355–368. doi:10.1016/S0734-189X(87)80186-X.
- [23] X. Wang, B. S. Wong, T. C. Guan, Image Enhancement for Radiography Inspection, in: Third International Conference on Experimental Mechanics and Third Conference of the Asian Committee on Experimental Mechanics, Vol. 5852, International Society for Optics and Photonics, 2005, pp. 462–469. doi:10.1117/12.621707. URL <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/5852/0000/Image-enhancement-for-radiography-inspection/10.1117/12.621707.short>
- [24] A. E. Hoerl, R. W. Kennard, Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics 12 (1) (1970) 55–67. doi:10.1080/00401706.1970.10488634. URL <https://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634>
- [25] Universidad Nacional de Colombia | Computer Imaging and Medical Ap-

- plications Laboratory.  
URL <http://cimalab.intec.co/?lang=en&mod=project&id=31>
- [26] U. R. Acharya, P. Chowriappa, H. Fujita, S. Bhat, S. Dua, J. E. Koh, L. W. Eugene, P. Kongmebhol, K. H. Ng, Thyroid lesion classification in 242 patient population using gabor transform features from high resolution ultrasound images, *Knowledge-Based Systems* 107 (2016) 235–245. doi:10.1016/j.knsys.2016.06.010.
- [27] U. R. Acharya, O. Faust, S. V. Sree, F. Molinari, J. S. Suri, Thyro-screen system: High resolution ultrasound thyroid image characterization into benign and malignant classes using novel combination of texture and discrete wavelet transform, *Computer Methods and Programs in Biomedicine* 107 (2) (2012) 233–241. doi:10.1016/j.cmpb.2011.10.001.
- [28] U. R. Acharya, S. V. Sree, M. R. K. Mookiah, F. Molinari, R. Garberoglio, J. S. Suri, Non-invasive automated 3d thyroid lesion classification in ultrasound: A class of thyroscan<sup>TM</sup> systems, *Ultrasonics* 52 (4) (2012) 508–520. doi:10.1016/j.ultras.2011.11.003.
- [29] J. Ding, H. Cheng, C. Ning, J. Huang, Y. Zhang, Quantitative measurement for thyroid cancer characterization based on elastography, *Journal of Ultrasound in Medicine* 30 (9) (2011) 1259–1266. doi:10.7863/jum.2011.30.9.1259.
- [30] U. Raghavendra, U. Rajendra Acharya, A. Gudigar, J. Hong Tan, H. Fujita, Y. Hagiwara, F. Molinari, P. Kongmebhol, K. Hoong Ng, Fusion of spatial gray level dependency and fractal texture features for the characterization of thyroid lesions, *Ultrasonics* 77 (2017) 110–120. doi:10.1016/j.ultras.2017.02.003.
- [31] U. Acharya, S. Sree, G. Swapna, S. Gupta, F. Molinari, R. Garberoglio, A. Witkowska, J. Suri, Effect of complex wavelet transform filter on thyroid tumor classification in three-dimensional ultrasound, *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 227 (3) (2013) 284–292. doi:10.1177/0954411912472422. URL <http://porto.polito.it/2512276/>
- [32] U. R. Acharya, O. Faust, S. V. Sree, F. Molinari, R. Garberoglio, J. S. Suri, Cost-effective and non-invasive automated benign & malignant thyroid lesion classification in 3d contrast-enhanced ultrasound using combination of wavelets and textures: a class of ThyroScan<sup>TM</sup> algorithms, *Technology in cancer research & treatment* 10 (4) (2011) 371–380. URL <http://journals.sagepub.com/doi/abs/10.7785/tcrt.2012.500214>
- [33] Y. Lecun, Y. Bengio, G. Hinton, Deep learning (2015). *arXiv:arXiv:1312.6184v5*, doi:10.1038/nature14539.
- [34] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, S. Sivaprasad, Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network, *Journal of Computational Science* 20 (2017) 70–79. doi:10.1016/j.jocs.2017.02.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S1877750317302028>
- [35] J. E. W. Koh, E. Y. K. Ng, S. V. Bhandary, A. Laude, U. R. Acharya, Automated detection of retinal health using PHOG and SURF features extracted from fundus images, *Applied Intelligence* doi:10.1007/s10489-017-1048-3. URL <https://doi.org/10.1007/s10489-017-1048-3>
- [36] J. Ma, F. Wu, T. Jiang, J. Zhu, D. Kong, Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images, *Medical Physics* 44 (5) (2017) 1678–1691. doi:10.1002/mp.12134. URL <http://doi.wiley.com/10.1002/mp.12134>
- [37] U. Raghavendra, H. Fujita, S. V. Bhandary, A. Gudigar, J. H. Tan, U. R. Acharya, Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images, *Information Sciences* 441 (2018) 41–49. doi:10.1016/j.ins.2018.01.051. URL <http://www.sciencedirect.com/science/article/pii/S0020025518300744>