Thyroid nodule classification using ultrasound elastography via linear discriminant analysis

Si Luo\(^a\), Eung-Hun Kim\(^b\), Manjiri Dighe\(^c\), Yongmin Kim\(^{a,b,*}\)

\(^a\)Department of Electrical Engineering, University of Washington, Seattle, WA 98195, United States
\(^b\)Department of Bioengineering, University of Washington, Seattle, WA 98195, United States
\(^c\)Department of Radiology, University of Washington, Seattle, WA 98195, United States

1. Introduction

A thyroid nodule is an abnormal growth of cells within the thyroid gland and can be non-cancerous (benign) or cancerous (malignant). Thyroid nodules are a common medical problem, with studies reporting as high as 50% of the population having a thyroid nodule at autopsy [1]. Thyroid nodules are typically asymptomatic and increasingly discovered incidentally during imaging examinations [2,3]. Although the majority of thyroid nodules are benign, it is clinically important to diagnose the small malignant population from the rest of the asymptomatic benign nodules. A fine needle aspiration (FNA) biopsy is used to evaluate a thyroid nodule’s malignancy and determine whether a surgical removal is warranted. It is estimated that somewhere between 250,000 and 300,000 thyroid FNA biopsies are performed annually in the United States. However, a large percentage (approximately 70%) of these biopsies turn out to be benign [4]. Thus, considering the increasing number of thyroid nodules being detected and the vast number of benign nodules undergoing FNA biopsies, the challenge lies in judiciously deciding which nodules should be aspirated [5].

Ultrasound (US) elastography measures the tissue deformation in response to stress to derive and display tissue stiffness [6]. Recent studies demonstrated the potential of applying US elastography to the thyroid gland in noninvasively differentiating between benign and malignant thyroid nodules [7–11]. In a study by Lyschik h et al. [12], the elastic modulus of excised thyroid tissues was shown to correlate with the malignancy of thyroid nodules. They observed that malignant thyroid nodules are five times stiffer than normal thyroid tissue while benign nodules are only 1.7 times stiffer than normal tissue.

Previous elastography studies [10,11,13] employed the free-hand external compression. Bae et al. [7] developed a new approach where the carotid artery was used as an in vivo compression source, taking advantage of its inherent periodic pulsation (e.g., expansion of the carotid artery lumen diameter during systole) and its position (adjacent to the thyroid). This approach was...
found to be advantageous over external freehand compression by
eliminating the interference caused by two independent compres-
sion sources on the thyroid (i.e., external freehand and carotid ar-
tery) [7].

To detect malignant nodules by thyroid elastography, the stiff-
ness of a nodule is mostly inferred by visually inspecting the pseudo-
color pattern in the thyroid strain image relative to the sur-
rounding tissues [9,11]. However, since the color pattern could var-
y depending on the amount of external freehand compression and
different observers could have different interpretations on the color pattern [14,15], the diagnosis based on this approach tends to be subjective. Rather than classifying thyroid nodules by visual inspection, Bae et al. [7,8] derived a thyroid stiffness index (TSI) to quantitatively evaluate the stiffness of a thyroid nodule. However, in utilizing TSI, one limitation was the manual placement of two regions of interest (ROIs) to estimate the strain near the carotid artery wall and in the thyroid nodule, which could introduce some variation.

In this paper, we present a thyroid nodule classification algo-
rithm using quantitative US elastography features.

2. Materials and methods

2.1. Case samples

Ninety-two patients (98 nodules), who were referred for an FNA biopsy following the Society of Radiologists in Ultrasound (SRU) guideline, were recruited for the study. The study was approved by the Institutional Review Board at our institution. Before enrollment, an informed consent was obtained from each participant. The mean age of patients was 52 ± 13 years (range 20–84 years, 72 females). The mean nodule size was 2.3 × 1.7 × 1.9 cm, ranging from 6.2 × 4.6 × 5.3 cm to 0.9 × 0.7 × 0.5 cm. US elastography was performed prior to the FNA procedure with a clinical ultrasound machine (Hi Vision 5500, Hitachi Medical Systems America, Twins-
burg, OH) with a 7.5-MHz linear array transducer. The data set included 82 benign and 16 malignant nodules. Unless a patient subsequently underwent surgery, we used FNA results as the diag-
nosis for thyroid nodules. For 19 patients who underwent surgery, the final diagnosis was based on the histopathological examination of the excised thyroid nodule (n = 22). One surgery patient had three papillary carcinomas, while another surgery patient had two benign nodules. Of the 22 excised nodules, 16 were diagnosed as papillary carcinoma, while 6 were diagnosed as benign.

2.2. Ultrasound elastography using carotid artery pulsation

Since the thyroid gland is adjacent to the carotid artery, the inherent periodic pulsation of the carotid artery was used for thy-
roid elastography [7]. Fig. 1 shows an example of thyroid strain rate waveform induced by pulsation of the carotid artery. The data were acquired from a healthy volunteer at 270 frames per second (fps). The strain was derived using the angular strain esti-
mator [16]. An ROI (a red rectangle1 in Fig. 1) was first placed
second (fps). The strain was derived using the angular strain esti-
data were acquired from a healthy volunteer at 270 frames per
rate waveform induced by pulsation of the carotid artery. The
roid elastography [7]. Fig. 1 shows an example of thyroid strain
inherent periodic pulsation of the carotid artery was used for thy-
nosis for thyroid nodules. For 19 patients who underwent surgery,
the final diagnosis was based on the histopathological examination
of the excised thyroid nodule (n = 22). One surgery patient had
three papillary carcinomas, while another surgery patient had
two benign nodules. Of the 22 excised nodules, 16 were diagnosed
as papillary carcinoma, while 6 were diagnosed as benign.

expansion in the axial direction during systole results in positive
peaks in Fig. 1.

2.3. Frequency characteristics of thyroid tissue deformation

The mechanical property of thyroid tissue can be modeled as a
system with a purely viscous damper and a purely elastic spring
connected in parallel [17]. The pulsation from the carotid artery
results in the periodic deformation observed in the thyroid tissue as
shown in Fig. 1. Because of the periodic nature of the input and
output, we can analyze the thyroid mechanical model in the fre-
quency domain. The carotid artery pulsation F(f) causes the thy-
roid tissue deformation X(f) via

\[ X(f) = H(f) \cdot F(f) \]

where H(f) represents the frequency response of thyroid tissue. The amount of deformation X(f) for benign and malignant nodules is ex-
pected to be different due to the difference in H(f). Fig. 2 shows the
strain rate waveform and its power spectrum for a benign nodule
and another set for a malignant nodule.

---

1 For interpretation of color in Fig. 1, the reader is referred to the web version
of this article.
For the strain rate waveforms in Fig. 2a and b, the peaks caused by the systolic blood pressure in the carotid artery lumen at the heart beat frequency (\( \sim 1 \text{ Hz} \)) are clearly visible. Because of the increased stiffness in a malignant nodule, its peak strain magnitude is smaller than that of a benign nodule. Another difference between benign and malignant nodules is the oscillation in the strain rate waveform. The strain rate waveform of the benign nodule (Fig. 2a) shows the noticeable oscillation, especially during the diastolic period, while the variation in the strain rate of the malignant nodule is minimum as shown in Fig. 2b. The oscillation frequencies are higher than the heart beat frequency as can be seen in the corresponding power spectrum (Fig. 2c and d).

According to the elastic spring model [17], the oscillation frequency is correlated to the natural frequency of thyroid tissue \( f_n \) by

\[
f_n = \frac{1}{2\pi} \sqrt{\frac{k}{m}}
\]

where \( m \) is the mass and \( k \) is the spring constant that is proportional to the thyroid tissue stiffness.

To analyze the frequency-dependent thyroid tissue deformation at different input frequencies, we model the carotid artery pulsation as

\[
F(t) = \sum_{f=f_h}^\infty C_f \cos(2\pi ft)
\]

where \( f_h \) is the heart beat frequency (\( \sim 1 \text{ Hz} \)) and \( C_f \) is the magnitude of compression at a specific frequency \( f \). Since the carotid artery is a low-frequency compression source, we model that the maximum magnitude \( C_{\text{max}} \) occurs at the heart beat frequency \( f_h \) and the compression magnitude decreases as the frequency increases. Then, the magnitude of thyroid tissue deformation can be calculated by [17]

\[
X(f) = \frac{C_f}{K} \frac{1}{\sqrt{(1 - r^2)^2 + (2\xi r)^2}}
\]

where \( \xi \) is the damping ratio and \( r \) is defined as

\[
r = \frac{f}{f_n}
\]

According to Eq. (4), various frequency components from the carotid artery pulsation would deform the tissue at their respective frequency, and the magnitude of deformation at a specific frequency is proportional to the ratio of input compression \( (C_f) \) and tissue stiffness \( (k) \). When \( f = f_h, r = \frac{f}{f_n} \ll 1 \). Thus, Eq. (4) can be simplified as
According to Eq. (6), the magnitude of thyroid tissue deformation at the heart beat frequency is inversely proportional to its spring constant or stiffness.

When the input frequency approaches the natural frequency of thyroid tissue, \(r\) approaches 1. At \(f_\text{h}\), the magnitude of deformation becomes

\[
X(f_h) = \frac{C_{max}}{k} \sqrt{\frac{m}{\nu}}
\]

(7)

by substituting \(\zeta = \frac{p}{r \nu m}\), into Eq. (4), where \(\nu\) is the viscosity of a nodule. According to Eq. (7), the mass (\(m\)) and viscosity (\(\nu\)) of a nodule play a role in determining \(X(f_h)\), but no quantitative data on \(m\) and \(\nu\) from benign and malignant thyroid nodules exist. Equation (7) also states that \(X(f_h)\) is dependent on the stiffness of tissue (\(k\)) and \(C_m\), which is the compression magnitude at the natural frequency. Due to the high stiffness (~3 times) of a malignant nodule, its natural frequency is higher than that of a benign nodule based on Eq. (2). Since \(C_T\) decreases as the frequency increases, the input component at the natural frequency to deform a malignant nodule is smaller than that of a benign nodule, which leads to smaller \(X(f_h)\). This could be a reason why the high frequency components of a malignant nodule in Fig. 2d are not prominent compared to Fig. 2c.

Because of the increased stiffness in malignant nodules, we can observe the frequency-domain characteristics in tissue deformation that are different from those of benign nodules. We can utilize this difference in performing the thyroid nodule classification. In the following sections, we present how to utilize these frequency characteristics in differentiating benign and malignant nodules.

2.4. Feature extraction and classification

We start the classification process by deriving the strain rate waveform and evaluating its frequency characteristics. The flowchart of our classification algorithm is shown by Fig. 3. Since a strain rate waveform represents the tissue deformation at a specific location, there are thousands of waveforms to be analyzed depending on the nodule size and parameters used in US scanning. The first step in the classification algorithm is to cluster together the waveforms with similar response to the carotid artery pulsation. We used k-means clustering [18] to perform this preprocessing with the following steps: (1) the boundary of a thyroid nodule is first delineated in the US B-mode image, and the segmented nodule is considered as a region of interest (ROI); (2) the strain rate waveforms from the ROI are organized into an \(n\) (rows) by \(p\) (columns) matrix, where \(n\) is the number of pixels or measurements within the ROI and \(p\) is the number of frames in the elastography data; (3) \(k\) rows are randomly selected from the matrix as the mean of each cluster, and \(k\) clusters are created by associating each measurement with the nearest mean based on the squared Euclidean distance; (4) the new mean of each of the \(k\) clusters is calculated; and (5) steps 3 and 4 are repeated until convergence is reached.

Fig. 4a and b shows the B-mode image of a thyroid nodule and the corresponding clustering results, where the pixels coded in the same color have the similar response to the carotid artery pulsation. The dashed area in Fig. 4a indicates the nodule ROI. In Fig. 4c, the blue line represents the strain rate waveform averaged over the region coded in blue in Fig. 4b while the red line corresponds to the red region in Fig. 4b. As can be seen, the mean strain for the red region is much lower than that for the blue region, which indicates that the red region is stiffer and more suspicious for malignancy. Fig. 4d shows the power spectrum of the mean strain rate waveform corresponding to the red color region. The power spectrum values at different frequencies as shown in Fig. 4d, ranging from 0 Hz to a half of the frame rate, were used as features in classification. If we denote a feature set as \(\Gamma\) with \(N\) elements, then a thyroid nodule can be considered as a point \(\Gamma\) in the \(N\)-dimensional feature space. For example, the power spectrum of a thyroid nodule with 128 bins can be considered as a vector of 128 dimensions or equivalently a point in a 128-dimensional space.

To perform classification directly in a high-dimensional space is difficult with the limited number of training cases (e.g., 98 nodules in our study) because many parameters need to be determined in a high-dimensional space compared to the number of parameters needed in a low-dimensional space. Typically, the high-dimensional feature space is converted into a space with fewer dimensions, where the classification can be performed more efficiently. Principal component analysis (PCA) is a technique used to approximate the original data with low-dimensional feature vectors [19]. The goal of PCA is to identify a set of orthogonal basis vectors for a new coordinate system to represent the original data set while preserving those characteristics that contribute the most to the data variance. By performing feature dimensionality reduction using PCA, the original \(N\)-dimensional feature vector \(\Gamma\) was transformed to a new feature vector \(\Omega\) with \(M (M < N)\) dimensions. This feature vector \(\Omega\) with \(M\) dimensions was used for nodule classification via linear discriminant analysis. In this study, we used \(N = 128\) and \(M = 15\).

Linear discriminant analysis (LDA) is a well-established pattern classification method. For a two-class problem, it determines a projection vector \(W\) to maximize the between-class scatter matrix while minimizing the within-class scatter matrix in the feature space [20]. The projected value of a nodule’s feature vector \(\Omega\) along \(W\) leads to a discriminant score for that nodule [21]. Since the feature vector \(\Omega\) is calculated from the power spectrum of the mean strain rate waveform of the most suspicious region, it represents the stiffness information of the stiffest part of a nodule. Smaller
spectral power in high frequency components of the feature vector leads to a larger discriminant score, which suggests the increased likelihood of malignancy.

The $k$-fold cross-validation method was used to verify the performance of our classification algorithm. Typically, $k$ was set to 10 [20]. All 98 data sets were partitioned into $k$ subsets, and the cross-validation process was repeated $k$ times. Each time, one of the $k$ subsets was retained as the validation set for testing the classifier, and the remaining $k-1$ subsets were used for training. The classification performance in terms of sensitivity and specificity in all $k$ trials was averaged.

3. Results

Table 1 summarizes the mean and standard deviation of discriminant scores given by LDA. The mean score of malignant nodules ($1.05 \pm 0.09, n = 16$) is significantly higher than that of benign nodules ($0 \pm 1.02, n = 82$) ($p = 0.0004$). The standard deviation of malignant nodule (0.09) was much smaller than that of benign nodules (1.02). The boxplot distribution of discriminant scores for benign and malignant nodules is shown in Fig. 5.

Fig. 6 shows an ROC curve to differentiate between benign and malignant nodules. The area under the ROC curve is 0.88. If we use a discriminant score of 0.86 as a threshold, we can obtain a sensitivity of 100% and specificity of 75.6% in detecting malignant nodules. Any nodule with the discriminant score less than 0.86 is classified as Type I (no FNA, observation-only), while any nodule with the score equal to or greater than 0.86 is classified as Type II (FNA). All of the malignant thyroid nodules were correctly classified, giving the sensitivity of 100% in detecting the malignant nodules. On the other hand, 20 benign thyroid nodules (five nodular goiters, 11 thyroid adenomas and four follicular lesions) were misclassified as Type-II nodules.

Previously, we used a program written in Matlab (MathWorks, MA) on a PC with an Intel Core 2 Dual 2.4-GHz processor and 3 Gbytes of memory to perform the elastography processing and nodule classification. It took 1.5 h to process one data set, which included strain estimation, $k$-means clustering and classification. Recently, we implemented our algorithm using C language and

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Malignant ($n = 16$)</th>
<th>Benign ($n = 82$)</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discriminant score</td>
<td>$1.05 \pm 0.09$</td>
<td>$0 \pm 1.02$</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Fig. 4. (a) B-mode image of a thyroid nodule (the carotid artery is indicated by a red arrow) and (b) the corresponding $k$-means clustering results, (c) the mean strain rate waveform of the blue and red regions in (b), and (d) the power spectrum of the mean strain rate waveform of the red region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
integrated it into a commercial ultrasound machine (Accuvix V20, Medison Co. Ltd., Korea), which has an Intel Core 2 Dual 2.4-GHz processor and 2 Gbytes of memory. Now, it takes less than 30 s, which enables the online analysis of patients’ data.

4. Discussion

Thyroid nodules are considered an epidemic due to the large number of imaging studies performed and the increasing incidental detection of these nodules [22]. Since the existing imaging modalities (CT, MRI and US) cannot accurately differentiate between the malignant and benign nodules, an FNA biopsy, which costs typically $1500 including an US exam, is performed on nodules showing suspicious features (e.g., size, microcalcifications and irregular margins). However, the majority of FNA procedures are performed on benign nodules. Thus, by detecting many benign nodules and removing them from an FNA procedure altogether, costs associated with FNA biopsies on patients with benign nodules could be substantially reduced.

Various ultrasound-based techniques to estimate the stiffness of tissue have been developed and shown the feasibility in differentiating malignant thyroid nodules from benign nodules. By applying external compression on the target area, the tissue strain could be derived by analyzing precompression and postcompression ultrasound signals. Rago et al. [11] showed the specificity and sensitivity as high as 100% and 97% in detecting the malignant thyroid nodules. However, the study subjects consisted of surgery-bound patients before thyroid removal rather than FNA-referred patients. In a recent study [9] with 145 thyroid nodules that were referred for surgical thyroid removal, a sensitivity of 88% and specificity of 90% with an accuracy of 88.9% were reported. For most current thyroid US elastography studies, the stiffness of a nodule is inferred by visually inspecting the pseudo-color pattern in the thyroid strain images relative to the surrounding tissues, after which a score, typically based on Ueno classification, is assigned [23]. However, the repeatability based on this approach is limited since the color pattern and the score could vary depending on the amount of externally-applied compression and multiple observers could have different interpretations on the color pattern [15,23]. Park et al. [14] reported no interobserver agreement for features observed on ultrasound elastography in evaluating malignant thyroid nodules.

In our previous study [7,8], we utilized the thyroid stiffness index (TSI) to quantify the stiffness of nodules. All the strain frames generated for this method (∼200) were averaged. From this single averaged frame, TSI was calculated as the ratio of the highest strain near the carotid artery to the lowest strain inside a nodule. A higher TSI value represents a stiffer thyroid nodule. By averaging, we simplified the TSI calculation, but the temporal information of a nodule’s response to the carotid artery pulsation was not utilized. During the current study, we have found that the strain rate waveform of a thyroid nodule is strongly correlated to its stiffness and this information could be used for nodule classification. Instead of directly using the strain rate waveform of a nodule as a feature, the power spectrum of the waveform was calculated and used in classification. By mapping the time-domain signal to the frequency domain, the information carried by the original signal was retained and represented as a uniform-length vector with the frequency range from 0 Hz to one half of the frame rate. Compared with averaging to get a single stiffness index, the frequency-domain feature can provide richer information as it retains more information on a nodule’s response to the carotid artery pulsation, which leads to better sensitivity and specificity in differentiating benign and malignant nodules compared to the TSI-based approach.

Compared to other thyroid elastography studies, our method can produce more quantitative and repeatable results because the stiffness of a nodule is scored by a classification algorithm rather than by an observer. For quite a few ultrasound elastography studies [9,11], the stiffness of a nodule was typically scored using 5 or 6 different grades based on the pseudo-color pattern...
in the elastography images. To make a diagnosis, the clinicians needed to visually inspect and categorize the pseudo-color pattern into one of these 5–6 different scores. Due to the subjective nature of this scoring method, the intra- and interobserver reliabilities need to be evaluated to assess the diagnostic consistency in the same observer and between observers. The kappa coefficient is the most commonly-used statistic for this purpose [24]. A kappa coefficient of 0 indicates chance agreement. In a recent prospective study with 193 breast lesions, Schaefer et al. [15] reported that the intra- and interobserver reliabilities for ultrasound elastography using the above scoring method are 0.720/0.561. The kappa coefficient of 0.561 corresponds to moderate agreement between observers [24], suggesting that 5–6 discrete scores might not be enough to account for the full variability in pseudo-color patterns for a nodule. In the current study, the stiffness of a nodule was calculated and used by a classification algorithm, where there was no need for further interpretation by an observer, leading to more consistent diagnostic results.

To use thyroid elastography as an FNA triage tool, it is important to capture as many malignant nodules as possible while excluding those nodules that are highly likely to be benign from FNA biopsies. As can be seen in boxplot distribution in Fig. 5, the malignant nodules get large discriminant scores with a small standard deviation. It indicates that we could exclude a large number of benign nodules that are located below the threshold of 0.86 without missing a malignant nodule. Thus, by utilizing US thyroid elastography as a triage tool, it would be possible to limit FNA biopsies to only Type-II (suspicious for malignancy) nodules, thereby increasing the percentage of malignant nodules being referred for an FNA biopsy from 16.3% to 44.4% and reducing the overall number of FNA biopsies by 63.3% (from 98 to 36).

In utilizing our classification method, one limitation is the contour delineation of a thyroid nodule, which may lead to somewhat different discriminant scores. Since the feature vector used for classification is derived from averaging strain rate waveforms within a cluster, which is typically composed of hundreds of pixels, the variability caused by the manual contour delineation should be minimized. To remove this variability, it is possible to incorporate an automatic or semi-automatic algorithm to segment a thyroid nodule in the US B-mode image [25]. Another possible limitation of the LDA classifier is that only linear combinations of the features are utilized. The performance of nonlinear classifiers (e.g., artificial neural network) could be evaluated and compared with that of the LDA classifier in the future. One issue of using LDA or other classifiers for thyroid nodule classification is that the classifier trained using the data sets acquired from one ultrasound machine may not be directly used in other machines. Also, parameter changes during scanning, such as transducer’s central frequency and frame rate, could also influence the classification results. However, this limitation could be overcome in the future by carefully evaluating the algorithm under various conditions, extracting optimal classification parameters for each setting and loading this information during the initialization step of thyroid elastography.

5. Conclusion

This study establishes the feasibility of using a linear discriminant classifier with ultrasound elastography features for thyroid nodule classification, which can be useful in FNA biopsy triage of thyroid nodules. Compared with other thyroid elastography methods, our classifier-based approach would give more objective and less observer-dependent results. By using our elastography method, we achieved the sensitivity of 100% and specificity of 75.6% for detecting malignant nodules in a retrospective study with 98 FNA-bound nodules, reducing the number of FNAs by 63.3%. Currently, we are planning to conduct a prospective study to confirm the efficacy of our elastography method as an FNA biopsy triage tool in managing thyroid nodules.

References