POLISH
JOURNAL *of* RADIOLOGY

Original paper

Differentiation between metastatic and tumour-free cervical lymph nodes in patients with papillary thyroid carcinoma by grey-scale sonographic texture analysis

Ali Abbasian Ardakani[1A,B,C,D,E,F], Alireza Rasekhi[2B,D,E,F], Afshin Mohammadi[2A,B,D,E,G], Ebrahim Motevalian[3B,D,E], Bahareh Khalili Najafabad[1C,E,F]

[1]Department of Medical Physics, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
[2]Department of Vascular and Interventional Radiology, Namazi Hospital, Shiraz University of Medical Sciences, Shiraz, Iran
[3]Department of Surgery, School of Medicine, Baqiyatallah Hospital, Baqiyatallah University of Medical Sciences, Tehran, Iran

Abstract

Purpose: Papillary thyroid carcinoma (PTC) is the most common thyroid cancer, and cervical lymph nodes (LNs) are the most common extrathyroid metastatic involvement. Early detection and reliable diagnosis of LNs can lead to improved cure rates and management costs. This study explored the potential of texture analysis for texture-based classification of tumour-free and metastatic cervical LNs of PTC in ultrasound imaging.

Material and methods: A total of 274 LNs (137 tumour-free and 137 metastatic) were explored using the texture analysis (TA) method. Up to 300 features were extracted for texture analysis in three normalisations (default, 3sigma, and 1-99%). Linear discriminant analysis was employed to transform raw data to lower-dimensional spaces and increase discriminative power. The features were classified by the first nearest neighbour classifier.

Results: Normalisation reflected improvement on the performance of the classifier; hence, the features under 3sigma normalisation schemes through FFPA (fusion Fisher plus the probability of classification error [POE] + average correlation coefficients [ACC]) features indicated high performance in classifying tumour-free and metastatic LNs with a sensitivity of 99.27%, specificity of 98.54%, accuracy of 98.90%, positive predictive value of 98.55%, and negative predictive value of 99.26%. The area under the receiver operating characteristic curve was 0.996.

Conclusions: TA was determined to be a reliable method with the potential for characterisation. This method can be applied by physicians to differentiate between tumour-free and metastatic LNs in patients with PTC in conventional ultrasound imaging.

Key words: computer-assisted, diagnosis, lymph nodes, pattern recognition, thyroid carcinoma, ultrasonography.

Introduction

Thyroid cancer is the most common endocrine malignancy, with approximately 64,300 new cases reported in the United States in 2016. The Surveillance, Epidemiology, and End Results (SEER) database reported that the incidence of thyroid cancer from 1975 to 2013 increased from 4.85 to 15.07 per 100,000 [1]. This increase was primarily a result of an increase in papillary thyroid carcinoma (PTC), the most common malignancy in thyroid cancer, representing approximately 90% of all thyroid malignancies [2,3]. PTC is the most common histologic type

of differentiated thyroid cancer and is characterised by metastasis to the cervical lymph nodes (LNs). LN metastasis occurs in approximately 30-90% of PTCs [4].

According to the American Thyroid Association (ATA), thyroid sonography of cervical LNs should be performed in patients with known or suspected thyroid nodules before patients undergo a thyroidectomy [5]. The ability to perform real-time scanning, its convenience, and the fact that it is radiation-free and inexpensive are advantages that make ultrasound examination a popular technique. The structure of LN metastasis is changed and differs from that of normal LN. Differences include cystic change, loss of normal architecture, calcification, and hyper-echogenicity [6]. These changes can affect the texture of an ultrasound image. An ultrasound image includes diverse grey-level intensities, and different tissues have different textures. Although there is no precise or mathematical definition of texture, it is simply conceived by the human eye. Image texture can be described by various patterns: coarse, fine, smooth, or spatial variations in pixel intensity of objects within an image. Structural abnormalities in an ultrasound image can be extracted by visual inspection, but complex patterns are difficult to interpret [7,8].

Biopsy is the best way to confirm whether an LN is involved or not; however, it is invasive, carries the risk of infection, and is expensive and time-consuming. The overall accuracy of LN fine-needle aspiration (FNA) was greater than 94% [9]. To avoid unnecessary biopsy, assuage anxiety, and increase diagnostic confidence, computerised texture analysis has been developed. Computerised texture analysis is a mathematical technique that detects pathological changes that cannot be perceived by the human eye, using conventional ultrasound imaging.

Recently, some computer-aided methods for characterising LN in ultrasound imaging have been proposed. Many quantitative features such as morphology [10-12], texture [13-15], contrast-enhanced based [16,17], and elastography [12,17-20] have been shown to be useful in differentiating between benign and malignant LNs. The present study provides additional information to support these claims. The current study evaluated texture ability, and this presents a non-invasive method to detect changes in ultrasound images of LN metastasis in patients with PTC. The variations of texture features between tumour-free (benign) and metastatic (malignant) LNs were used for classification.

Material and methods

Patients and image acquisition

Ultrasound images of 274 patients, comprising 137 tumour-free and 137 metastatic cervical LNs with proven FNA, were used in this retrospective study. Patient consent was not needed because the study was retrospective

and used archived data. All patients had a previous history of total thyroidectomy due to PTC. The ultrasonographic inclusion criteria for lymph node biopsy were as follows: irregular cystic changes, microcalcified lymph node, focal or diffuse hyperechogenicity, a length greater than 5 mm in the short axis diameter, round shape (long-to-short axial diameter ratio < 1.5), and loss of fatty echogenic hilum. The exclusion criteria were nondiagnostic biopsy and refusal to repeat FNAB, atypical/inconclusive FNAB and those who had prior head and neck irradiation or oncological surgery.

Neck ultrasonography was performed on these patients as routine evaluation. The largest cervical LN was selected because of the greater possibility of it exhibiting malignancy, and FNA cytology was performed to confirm or exclude involvement. Patients with no diagnostic or atypical/inconclusive FNA results were not included in the study.

All FNA were performed after ultrasound images of the target LN had been obtained for FNA under guidance of ultrasonography. Ultrasonography was performed using the Accuvix V20 sonography system (Medison, Seoul, Korea) equipped with an L5-13IS (5-13 MHz) linear array transducer. Ultrasonography of the cervical triangle was performed, and all detected LNs were evaluated in longitudinal and transverse sections to measure the largest small diameter and to select the biggest and target LN for FNA. Only one image per patient with the maximum longitudinal section of each node was used for texture analysis.

Texture analysis

MaZda software version 4.6 (Technical University of Lodz, Institute of Electronics) was used for texture analysis. One 2D region of interest (ROI) was selected for each node; hence, 274 non-overlap ROIs (137 tumour-free and 137 metastatic) were selected for discrimination and classification. The sizes of the ROIs depended on nodes, where the mean size of the ROI was 15×15 mm. Either an endocrinologist or a general surgeon referred the cases based on clinical indication and routine patient follow-up. The images were compared with FNA cytology and the reference for statistical analysis.

Texture features

Up to 300 texture features were recorded based on histogram, absolute gradient (spatial variation of grey-level values), run-length matrix (counts of pixel runs with the specified grey-scale value and length in a given direction), co-occurrence matrix (information about the distribution of pixel pairs separated by a given distance and direction), auto-regressive model (description of correlation between neighbouring pixels), and wavelets (decomposition image frequencies in different scales).

Normalisation schemes

In this study, texture features were calculated for each ROI in three normalisations. The first, N_1, was the default, and no normalisation was applied. Thus, the images had the same appearance at an intensity range of 1 to 2^k, where k is the number of bits per pixel. The second, N_2, was 3sigma, in which the ROI intensities were limited in the range [$\mu - 3\sigma$ and $\mu + 3\sigma$], where μ and σ, respectively, are the mean value and standard deviation of the intensity. The third normalisation was N_3, 1-99%, in which the image intensity ranges were normalised between the darkness (accumulated histogram is equal to 1% of total) and brightness levels (accumulated histogram = 99% of total) inside the ROI. The intensity levels outside the normalisation range were not considered for analysis.

Automated feature selection

None of the 300 texture features were suitable or effective for differentiating between tumour-free and metastatic LNs. By using two feature-reduction algorithms, Fisher, and the lowest probability of classification error and average correlation coefficients (POE+ACC), the parameters were reduced to the best ten texture features to show the best discrimination between tumour-free and metastatic LNs [21,22]. Also, the fusion Fisher plus the POE+ACC (FFPA) were performed to examine the discriminatory power of the two algorithms together.

On Fisher algorithm up to ten features were selected with the highest Fisher coefficient, defined as the ratio of between-class variance (D) (computed between the class means μ_k, k = 1, 2, K) to within-class variance (V) (computed between the samples of class k and the corresponding class mean μ_k):

$$F = \frac{D}{V} = \frac{\frac{1}{1 - \Sigma_{k=1}^{K} P_k^2} \Sigma_{k=1}^{K} \Sigma_{j=1}^{K} P_k P_j (\mu_k - \mu_j)^2}{\Sigma_{k=1}^{K} P_k V_k} \tag{1}$$

where μ_k and V_k are mean and variance of class k, F is Fisher coefficient, and P_k is probability of class k. Maximisation of the Fisher coefficient is desirable.

A POE+ACC algorithm uses the measures of both probability of classification error (POE) and average correlation coefficients (ACC) between chosen features to introduce 10 features with the lowest POE+ACC. The first feature, f_i, is selected to minimise POE for all classes:

$$f^1 = f_i : \min_i [POE (f_i)] \tag{2}$$

where POE_{fi} (is the classification error probability for feature f_i. This probability is defined as a ratio of a number of corrections in a sample classification to the whole number of samples in the analysed data set using only feature f_i for classification.

$$POE(f_i) = \frac{\text{number of samples not properly classified;}}{\text{total number of samples}} \atop \text{in analysed data set} \tag{3}$$

Figure 1 shows the sample distribution of feature f_i for two classes. The samples marked can be assigned to either class 1 or 2. In this case, these samples cannot be properly classified. The next feature (second feature) is then selected by minimising the sum of all features except f_i:

$$f^2 = f_i : \min_i [POE(f_i) + |CC (f^1 \cdot f_i)|] \tag{4}$$

The $|CC(f^1 \cdot f_i)|$ is the absolute value of the correlation coefficient between the previously calculated feature f_i and the new feature f_i. The nth feature is selected by minimising the following sum for all remaining features excluding already chosen features:

$$f^2 = f_i : \min_i [POE (f_i) + \frac{1}{n-1} \Sigma_{k=1}^{n-1} |CC (f^k \cdot f_i)|] =$$
$$= \min_i [POE(f_i) + ACC(f_i)] \tag{5}$$

where the average sum is extended for all correlation coefficients between previously selected features and feature f_i. This sum is defined as an average correlation coefficient (ACC).

In brief, the POE+ACC algorithm introduces features with high discriminatory potential and the least correlation with features that are already selected. The Fisher algorithm selected features with maximised differences between two groups. Also, the FFPA algorithm introduce features with the lowest POE+ACC and highest difference between two groups.

Statistical analysis and classification

Linear discriminant analysis (LDA) was used to transform raw texture features to lower-dimensional spaces and to increase discriminative power; LDA seeks the most efficient direction for maximal separation of features. LDA demonstrated that variability among feature vectors
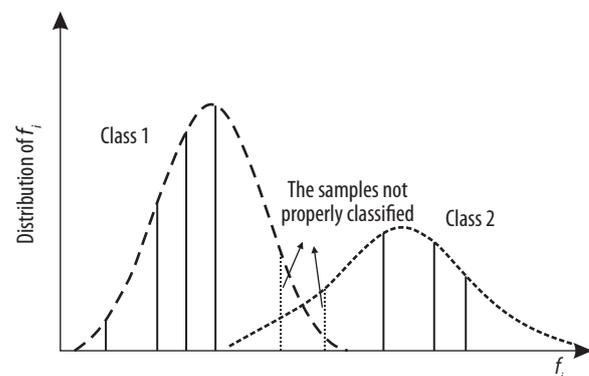
Figure 1. A sample distribution of feature for two classes. The samples marked cannot be properly classified

of the same class (within class scatter) was minimised and variability among the feature vectors of different classes (between class scatter) was maximised. Features processed by LDA were considered useful for pattern recognition and classification because they put data of the same class closer together and data of different classes further apart.

The first nearest neighbour (1-NN) classifier was used for the features resulting from LDA. To compare diagnostic performances, five well-known indexes were calculated: accuracy (ACC), sensitivity (SEN), specificity (SPC), positive predictive value (PPV), and negative predictive value (NPV). Their definitions are given as:

$$\text{Accuracy (ACC)} = \frac{N_{TN} + N_{TP}}{N_{TN} + N_{FN} + N_{TP} + N_{FP}} \quad (6)$$

$$\text{Sensitivity (SEN)} = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (7)$$

$$\text{Sensitivity (SPC)} = \frac{N_{TN}}{N_{TN} + N_{FP}} \quad (8)$$

$$\text{Positive predictive value (PPV)} = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (9)$$

$$\text{Positive predictive value (NPV)} = \frac{N_{TN}}{N_{TN} + N_{FN}} \quad (10)$$

where N_{TP} and N_{TN} are the numbers of correctly diagnosed metastatic and tumour-free cases, respectively. N_{FP} and N_{FN} are the numbers of incorrectly diagnosed metastatic and tumour-free cases, respectively.

The area under the ROC curve (A_z) was also calculated to evaluate the overall performance of the proposed method [23]. ROC analysis was performed using SPSS software version 19 (SPSS Inc., Chicago, USA), and A_z values were estimated to be beyond the 95% confidence level. Figure 2 shows the steps of the texture analysis process.
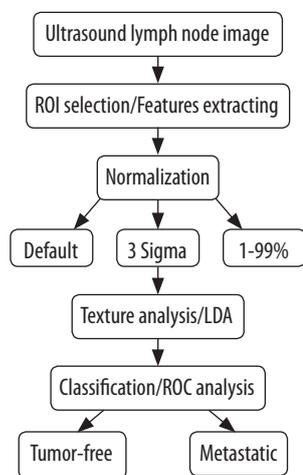

Figure 2. Overview of general texture analysis process in the ultrasound lymph node images

Results

Fisher and POE+ACC features

Figure 3A and Table 1 show the 10 best features with the highest Fisher coefficient values. Wavelet and run-length matrix features were more predominant and effective features for discriminating tumour-free from metastatic LNs in all three normalisations. The Fisher features are mostly from the energy of wavelet coefficient in "low-high" energy components in second level wavelet decomposition (WavEnLH_s-2), WavEnLH_s-3, WavEnLH_s-4, and WavEnLH_s-5, and "low-low" energy components in first-level wavelet decomposition (WavEnLL_s-1), WavEnLL_s-2, WavEnLL_s-3, WavEnLL_s-4, and WavEnLL_s-5 from wavelet; run-length non-uniformity in horizontal and 45-degree directions (Horzl_RLNU and 45dgr-RLNU), grey-level non-uniformity of image in horizontal, vertical, 45-degree, and 135-degree directions (Horz_GLNU, Vert_GLNU, 45dgr_GLNU, and 135dgr_GLNU) from the Run-length matrix; sum average S(2, 2) (SA_S[2, 2]) and sum entropy S(0, 5) (SE_S[0, 5]) from the co-occurrence matrix where $S_{(i,j)}$ shows the direction of matrix construction and interpixel distance i along rows and j along columns of the matrix, and skewness of absolute gradient (Gr_Skewness) from the gradient.

Figure 3B and Table 2 show the 10 best features with the lowest POE+ACC values. Despite Fisher, only the wavelet features in POE+ACC were predominant and effective. Many common parameters could be seen between Fisher and POE+ACC reduction feature methods. The other POE+ACC features were mostly from the energy of the wavelet coefficient in "high-high" energy components in fourth- and fifth-level wavelet decomposition (WavEnHH_s-4 and WavEnHH_s-5) from wavelet, long run emphasis in a horizontal direction (Horz_LRE), and fraction in a vertical direction (Vert_Fraction) from the run-length matrix; Teta4 and Teta3 from the autoregressive model; percentile 90% (Perc_90%), variance, and skewness from histogram; kurtosis of absolute gradient (Gr_Kurtosis) from Gradient and sum average S(0, 5) (SA_S[0, 5]), correlation S(0, 5) (Correlat_S[0, 5]), and sum of squares S(4, −4) (SOS_S[4, −4]) from the co-occurrence matrix.

Texture classification

The diagnostic performance of the texture analysis methods and normalisation schemes for Fisher features are listed in Table 3. Considering the normalisation role, Fisher features represented a higher performance with 93.43% sensitivity, 91.97% specificity, 92.70% accuracy, 92.75% PPV, and 93.33% NPV in 3sigma, while classification tasks carried out in 1-99% and default normalisation had subsequent ranks, respectively.

POE+ACC features have a higher discrimination power to distinguish between tumor-free and metastatic
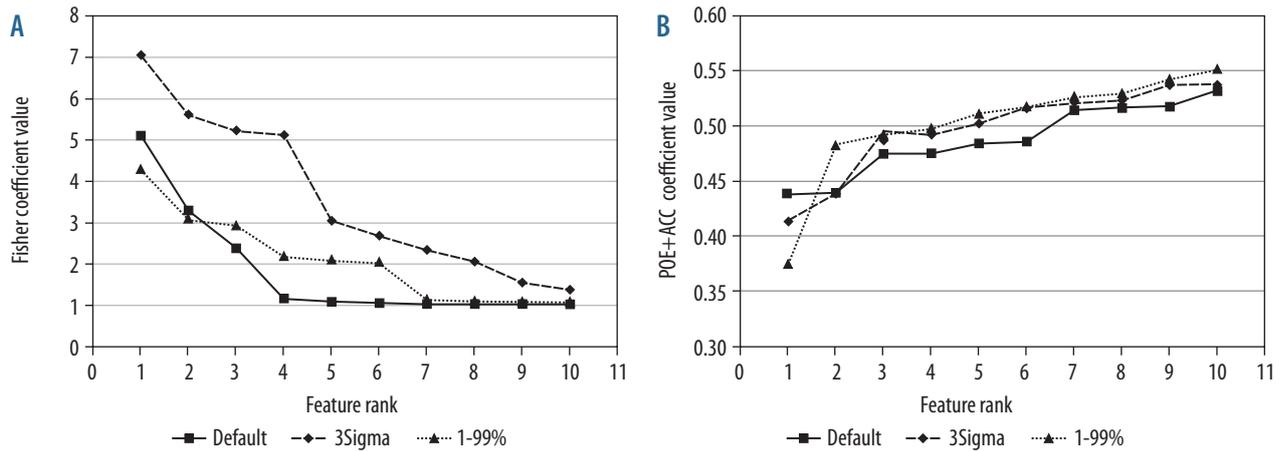
Figure 3. Evolution of two reduction methods for texture analysis. **A)** Fisher coefficient with 10 highest values, (**B**) probability of classification error and average correlation coefficient (POE+ACC) with 10 lowest values

Table 1. Summary of 10 best Fisher features with highest values

Feature rank	Normalisation					
	Default		3sigma		1-99%	
	Feature	Fisher coefficient	Feature	Fisher coefficient	Feature	Fisher coefficient
1	WavEnLH_s-4	5.1055	WavEnLL_s-3	7.0694	WavEnLH_s-4	4.2902
2	WavEnLH_s-3	3.3025	WavEnLL_s-2	5.6225	WavEnLH_s-3	3.1058
3	WavEnLH_s-5	2.3952	WavEnLL_s-4	5.2312	WavEnLH_s-5	2.9345
4	WavEnLH_s-2	1.1781	WavEnLH_s-4	5.1331	Gr_Skewness	2.1927
5	WavEnLL_s-5	1.0984	WavEnLH_s-3	3.0562	Horzl_RLNU	2.1008
6	WavEnLL_s-4	1.0715	WavEnLL_s-1	2.6987	Horzl_GLNU	2.0388
7	Horzl_GLNU	1.0465	WavEnLL_s-5	2.3642	45dgr_RLNU	1.1473
8	45dgr_GLNU	1.0410	WavEnLH_s-5	2.0884	45dgr_GLNU	1.1190
9	135dr_GLNU	1.0376	SE_S(0,5)	1.5735	135dr_GLNU	1.1096
10	Vertl_GLNU	1.0371	SA_S(2,2)	1.4091	Vertl_GLNU	1.1036

Table 2. Summary of 10 best probability of classification error and average correlation coefficient (POE+ACC) features with lowest values

Feature rank	Normalisation					
	Default		3sigma		1-99%	
	Feature	POE+ACC coefficient	Feature	POE+ACC coefficient	Feature	POE+ACC coefficient
1	WavEnLH_s-4	0.4380	WavEnLH_s-4	0.4124	WavEnLH_s-4	0.3759
2	Horz_LRE	0.4387	SA_S(0,5)	0.4377	Vertl_Fraction	0.4820
3	WavEnLH_s-5	0.4737	WavEnLH_s-3	0.4868	Horzl_GLNU	0.4918
4	45dgr_GLNU	0.4745	WavEnHH_s-4	0.4913	WavEnLH_s-3	0.4967
5	WavEnLL_s-4	0.4831	WavEnLL_s-3	0.5013	WavEnLH_s-5	0.5099
6	WavEnLH_s-3	0.4845	Teta4	0.5158	SOS_S(4,-4)	0.5169
7	WavEnLH_s-2	0.5142	Horzl_GLNU	0.5210	WavEnHH_s-5	0.5266
8	WavEnHH_s-4	0.5166	Variance	0.5220	Perc.90%	0.5280
9	Correlat_S(0, 5)	0.5176	WavEnLH_s-5	0.5372	Teta3	0.5414
10	Skewness	0.5325	Gr_Kurtosis	0.5373	WavEnLL_s-5	0.5514

Table 3. Summary of performance for Fisher feature reduction method in three normalisation schemes

Normalisation scheme	SEN (%)	SPC (%)	ACC (%)	PPV (%)	NPV (%)	A_z value*	Correct classification
Default	89.78	87.59	88.69	87.88	89.55	0.908 (0.872, 0.944)	243/274 (88.68%)
3sigma	93.43	91.97	92.70	92.75	93.33	0.943 (0.917, 0.970)	254/274 (92.70%)
1-99%	89.78	91.24	90.51	91.11	89.93	0.925 (0.893, 0.956)	248/274 (90.51%)

SEN – sensitivity, SPC – specificity, ACC – accuracy, PPV – positive predictive value, NPV – negative predictive value, A_z – area under ROC curve
*Numbers in parentheses are 95% confidence intervals

Table 4. Summary of performance for probability of classification error and average correlation coefficient (POE+ACC) feature reduction method in three normalisation schemes

Normalisation scheme	SEN (%)	SPC (%)	ACC (%)	PPV (%)	NPV (%)	A_z value*	Correct classification
Default	90.51	90.51	90.51	90.51	90.51	0.927 (0897, 0.958)	248/274 (90.51%)
3sigma	91.97	91.24	91.60	91.30	91.91	0.934 (0.906, 0.963)	251/274 (91.60%)
1-99%	95.62	93.43	94.52	93.57	95.52	0.963 (0.943, 0.982)	259/274 (94.52%)

SEN – sensitivity, SPC – specificity, ACC – accuracy, PPV – positive predictive value, NPV – negative predictive value, A_z – area under ROC curve
*Numbers in parentheses are 95% confidence intervals

Table 5. Summary of performance for fusion Fisher and probability of classification error and average correlation coefficient (POE+ACC) feature reduction method in three normalisation schemes

Normalization scheme	SEN (%)	SPC (%)	ACC (%)	PPV (%)	NPV (%)	A_z value*	Correct classification
Default	92.70	92.70	92.70	92.70	92.70	0.945 (0.917, 0.972)	254/274 (92.70%)
3sigma	99.27	98.54	98.90	98.55	99.26	0.996 (0.990, 1.000)	271/274 (98.90%)
1-99%	96.35	97.08	96.71	97.06	96.38	0.979 (0.965, 0.994)	265/274 (96.71%)

SEN – sensitivity, SPC – specificity, ACC – accuracy, PPV – positive predictive value, NPV – negative predictive value, A_z – area under ROC curve
*Numbers in parentheses are 95% confidence intervals

LNs, in which sensitivity, specificity, accuracy, PPV, and NPV were 95.62%, 93.43%, 94.52%, 93.57%, and 95.52%, respectively, in 1-99% normalization, while classification tasks carried out in 3sigma and default normalisations were in the subsequent ranks, respectively (Table 4).

The FPPA features achieved discriminating power with a sensitivity of 99.27%, specificity of 98.54%, accuracy of 98.90%, PPV of 98.55%, and NPV of 99.26% in 3sigma normalisation; 1-99% and default normalisation were in subsequent ranks, respectively (Table 5).

Figure 4 represents the ROC curves of texture analysis on the same graph for each feature reduction algorithm to compare the discriminating power of classification based on normalisation schemes. 3sigma normalisation has an advantage over the default and 1-99% normalisation in Fisher and FFPA features because of a greater A_z value. As shown, the best performance was obtained in Fisher and FFPA features with A_z values of 0.943 and 0.996, respectively. In POE+ACC features, the best accuracy was achieved at 1-99% normalisation with an A_z value of 0.963.

The data distributions of the two texture classes with the best results in FFPA features and each normalisation are illustrated in Figure 5. As can be seen, the greatest power of discrimination to distinguish between tumour-free and metastatic LNs was achieved in FFPA features and 3sigma normalisation schemes.

Discussion

Discriminating between tumour-free and metastatic LNs is one of the most critical factors for improving the accuracy of radiologists' initial diagnoses. The primary objective of this study was to evaluate texture ability, a non-invasive method for identifying changes between tumour-free and metastatic LNs using ultrasound imaging. The results of the current study proved that texture analysis can differentiate between tumour-free and metastatic LNs with high accuracy.

Two options for texture analysis were used and provided a total of nine states for each ROI case study: three feature-reduction methods (Fisher, POE+ACC, and FFPA) and three normalisation schemes (default, 3sigma, and 1-99% normalisation). The results demonstrated that FFPA features had more discriminative power than Fisher and POE+ACC features in the data of this research, where the best results were driven in FFPA features and 3sigma normalisation, with an value of 0.996, which corresponded to a sensitivity of 99.27%, specificity of 98.54%, accuracy of 98.90%, PPV of 98.55%, and NPV of 99.26%.

In all normalisation schemes in order to A_z values, FFPA features had a higher performance than the other two algorithms, i.e. Fisher and POE+ACC. The A_z values in each normalisation were as follows: default (Fisher vs. POE+ACC

vs. FFPA: 0.908 vs. 0.927 vs. 0.945), 3sigma (0.943 vs. 0.934 vs. 0.996), and 1-99% normalisation (0.925 vs. 0.963 vs. 0.979) (Figure 4). The Fisher method selected 10 features based on the maximised variance (difference) between textures, and POE+ACC minimised the number of sample classification errors. Because the criteria for feature selection by Fisher and POE+ACC are different, it can be concluded that these two methods are complementary. Hence, the fusion of these two methods showed improved performance on the classification task. From Tables 3-5 and Figure 4, it can be seen that normalisation reflected improvement in the performance of the classifier. Therefore, the best performance in this study was obtained in 3-sigma normalisation with FFPA features.

In the last decade, computer-aided diagnosis has been employed to classify LNs using ultrasound imaging. Ultrasound imaging includes many kinds of features, such as textural, elastographic, and morphological features, which are useful for classification tasks. Elastographic features contain information about tissue stiffness. Ultrasound elastography is a non-invasive technique used to achieve tissue deformation in response to compression [24]. Studies have used ultrasound elastography to classify benign and malignant LNs. Azizi et al. [18] used shear-wave elastography (SWE) to evaluate stiffness in cervical LN by measuring shear-wave velocity (SWV) on ultrasound images. They achieved a sensitivity of 92.52%, specificity of 75.46%, PPV of 48.54%, NPV of 97.60%, and of 0.88 in classifying benign and malignant cervical LNs. Cheng et al. [20] achieved a sensitivity of 78.9%, specificity of 74.4%, accuracy of 77%, and A_z of 0.855 using SWV measurements. Park et al. [19] combined SWE and ultrasound imaging to distinguish benign from malignant cervical LN metastases in PTC with a sensitivity of 44.6%, specificity of 87.1%, PPV of 65.9, NPV of 73.8%, and of 0.667 for central LNs and a sensitivity of 95.8%, specificity of 58.4%, PPV of 14%, NPV of 99.5%, and A_z of 0.924 for lateral LNs. In a study by Acu et al. [25], the strain index from ultrasound images classified benign and malignant cervical LNs with a sensitivity, specificity, and accuracy of 71.6%, 76.5%, and 75.0%, respectively.

Morphological features describe margin irregularity, symmetry, and shape of the tumour surface. In this regard, Chmielewski et al. [10] extracted morphological parameters from normalised signed distance (NSD) transforms to classify benign and malignant axillary LNs and achieved a sensitivity and specificity of 90% and an A_z of 0.95. Jeong et al. [11] showed that the Yonsei Estimated Value (YEV) scoring system is useful; they achieved a performance in which sensitivity, specificity, PPV, and NPV were 76.3%, 69.8%, 56.7%, and 85%, respectively. Desmots et al. [12] extracted morphological features from ultrasound images and classified cervical LNs with a sensitivity of 80%, specificity of 94%, PPV of 92%, NPV of 83%, and A_z of 0.885. Acu et al. [25] indicated that morphological features can classify benign and malignant LNs with 97% sensitivity, 31.4% specificity, and 51.3% accuracy.
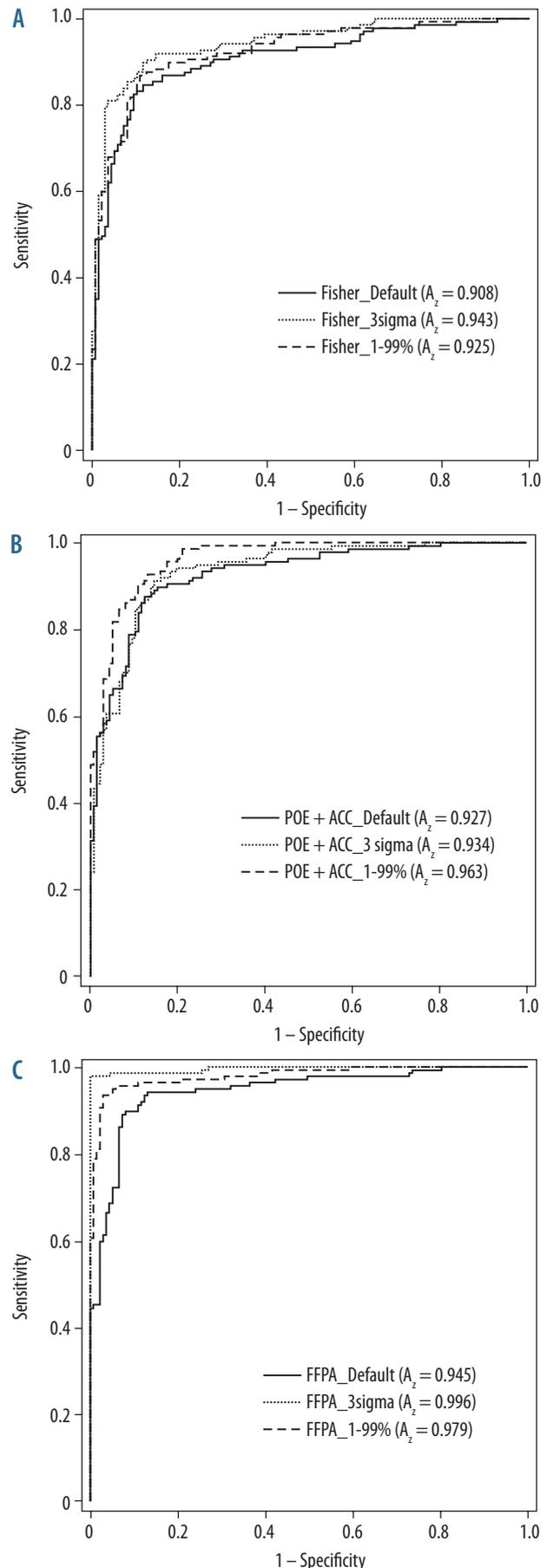

Figure 4. The diagrams of the ROC curves for texture analysis method in Fisher (**A**), POE+ACC (**B**), and FFPA feature reduction method (**C**)
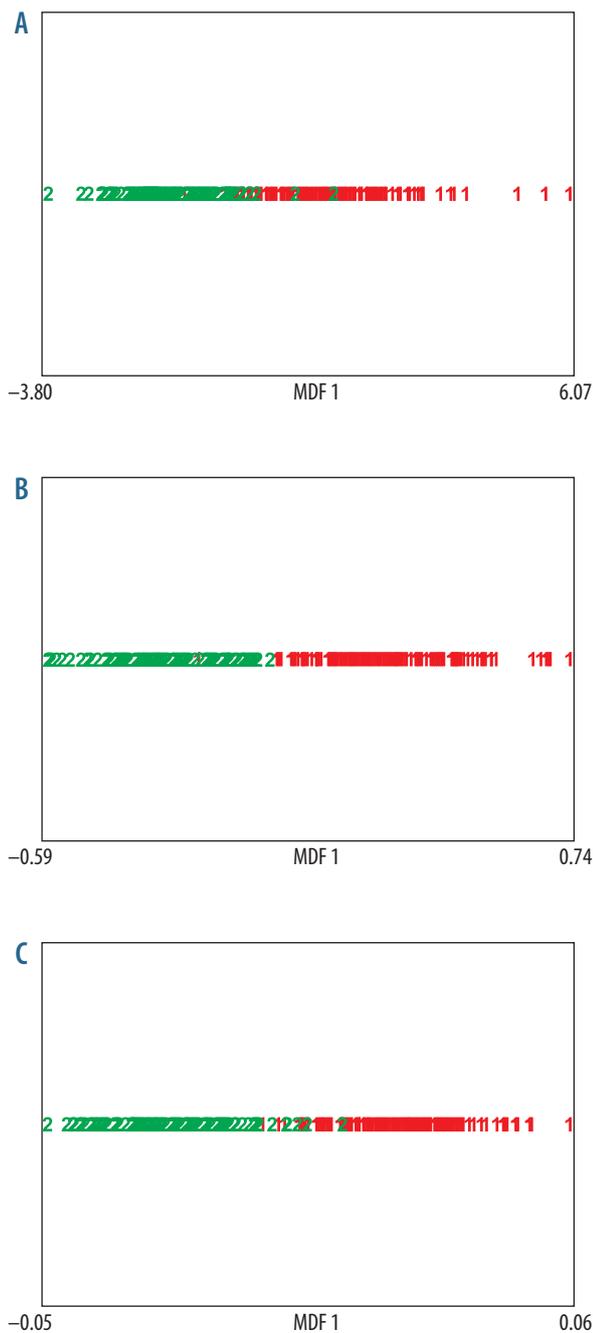
Figure 5. Sample distributions after texture analysis with FFPA features. **A**) default, (**B**) 3sigma, and (**C**) 1-99% normalisation. MDF – most discriminating features; "1" and "2" represent tumour-free and metastatic lymph nodes, respectively

In texture analysis methods, Chen *et al.* [13] used a combination of texture features based on the co-occurrence matrix and neighbouring grey-level dependence matrix (NGLDM) to classify benign and malignant LNs on ultrasound images. They achieved a 100% classification accuracy using a support vector machine classifier. In another study, Zhang *et al.* [14], used features based on the co-occurrence matrix and presented a good classification of benign and malignant cervical LNs with an A_z of 0.827. Chang *et al.* [15] utilised seven texture feature subsets to

characterise LNs in ultrasound images. They subdivided LNs into six groups and reached average accuracy, sensitivity, and specificity rates of 99.10%, 96.43%, and 99.53%, respectively, in the classification task.

Another technique that can be used to differentiate benign from malignant LNs is contrast-enhanced ultrasound (CEU). Dudau *et al.* [16] indicated that CEU was useful for classifying LNs. They classified benign and malignant LNs with a sensitivity of 100% and specificity of 85.7%. Rubaltelli *et al.* [26] indicated that CEU showed 100% sensitivity and 99.5% specificity in classifying LNs.

In some studies, features from different groups were combined to achieve the best performance. Zhang *et al.* [14] combined texture and morphological features from ultrasound images and indicated that the combined features achieved better results in classifying LNs with an A_z of 0.892 in comparison with texture alone ($A_z = 0.827$). However, in a study by Acu *et al.* [25], elastographic features had no additional value regarding the performance of morphological features.

In the current study, texture features from six main categories were analysed to differentiate between tumour-free (benign) and metastatic (malignant) LNs. These features presented a good classification with a sensitivity of 99.27%, specificity of 98.54%, accuracy of 98.90%, PPV of 98.55%, and NPV of 99.26%. The area under the receiver operating characteristic curve was 0.996. This means that about 99% of patients with uncertain malignant or benign LNs can avoid unnecessary FNA if the texture analysis method introduced here is used to diagnose these patients. In comparison with the other texture analysis studies, we can achieve a higher accuracy in classification of tumour-free and metastatic cervical LNs [13-15]. All of them focused predominantly on co-occurrence matrix features. But this study implies additionally that wavelet-based texture features play an important role in distinguishing tumour-free from metastatic cervical LNs. Wavelet has an advantage over other texture features groups die to has an information in both spatial and frequency domains.

Three normalisations and three feature-reduction algorithms were employed and provided a total of six states per ROI. Each set was examined individually to determine the best features descriptor for differentiation between tumour-free and metastatic LNs. Based on the authors' hypothesis, such conditions (under FFPA features and 3sigma normalisation) will classify groups with more confidence, and this was shown by the results. The results of the current study also indicate that texture features possess significantly greater discriminative ability than other features, such as morphology-, CEU-, and elastography-based features.

Although some effort has been made to investigate correlations between CEU features and pathological changes, the findings are quite controversial in the literature. In this regard, Dudau *et al.* [16] and Rubaltelli *et al.* [26] indicated that CEU features were promising in dif-

ferentiating between benign and malignant LNs, but Tan *et al.* [17] achieved the opposite result. Furthermore, for Doppler, the resistive index (RI) and vascular index (VI) were not significantly correlated to LN metastasis [12,27]. In addition, Acu *et al.* [25] indicated that the accuracy of B-mode and Doppler sonography was higher than that of elastography, and elastography adds no additional information. Although Doppler, CEU, and elastography are still used in many centres, there remains the need for further research in this field. Despite all this, all previous studies have indicated that texture features have a higher accuracy than other sonographic features, and the present study confirms this.

Since the physician-referred cases were based on clinical indications, other node characteristic features had no additional information and were not considered for TA and classification. This method is not suggested as an alternative to biopsy, but it can be applied to help radiologists identify subtle changes and select patients with a high risk of malignant LNs for biopsy.

A previous study by the current authors indicated that texture analysis based on wavelet is a useful tool for discriminating between benign and malignant thyroid nodules, with 100% overall accuracy in ultrasound imaging [28]. The present study also proved that wavelet texture features have a high ability to classify LNs. The methods used in this study can aid radiologists in distinguishing between tumour-free and metastatic LNs in the target nodes and it can improve the overall accuracy of radiologists from an A_z value of 0.865 to 0.900 [29]. The main advantage of this method is that there is no operator dependency because the analysis is performed by the computer. Moreover, it requires no additional time or costs.

There were several limitations in the current study. First, in the MaZda software, combination tools were not available. This made it difficult to perform some calculations. For example, the averaging of run-length matrix features of four different orientations was difficult. Second, no diagnostics from radiologists were implemented in this study. The texture analysis results were compared only with pathology. Further investigation comparing the texture analysis results with radiologist diagnostics to evaluate the radiologists' performance is needed. Third, the sonographic classification of the current study was compared with FNA biopsy results. Although FNA biopsy is highly sensitive for the diagnosis of malignancy, it cannot replace surgical biopsy. Hence, definitive results require surgical biopsy analysis.

Conclusions

In conclusion, a new approach based on texture analysis is proposed for the evaluation of LN metastasis 2-dimensional sonography. Preliminary results indicate that texture features of a conventional ultrasound image can be used as a supplementary technique to improve the radiologist's understanding of conventional ultrasound imaging in characterising LNs.

Conflict of interest

The authors report no conflict of interest.

References

1. Howlader N, Noone A, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2013. Secondary SEER Cancer Statistics Review, 1975-2013 based on November 2015 SEER data submission, posted to the SEER web site, April 2016; https://seer.cancer.gov/csr/1975_2013/.

2. Noone AM, Cronin KA, Altekruse SF, et al. Cancer incidence and survival trends by subtype using data from the Surveillance Epidemiology and End Results Program, 1992-2013. Cancer Epidemiol Biomarkers Prev 2017; 26: 632-641

3. Davies L, Welch HG. Current thyroid cancer trends in the United States. JAMA Otolaryngol Head Neck Surg 2014; 140: 317-322.

4. Choi YJ, Yun JS, Kook SH, et al. Clinical and imaging assessment of cervical lymph node metastasis in papillary thyroid carcinomas. World J Surg 2010; 34: 1494-1499.

5. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2016; 26: 1-133.

6. Nachiappan AC, Metwalli ZA, Hailey BS, et al. The thyroid: review of imaging features and biopsy techniques with radiologic-pathologic correlation. Radiographics 2014; 34: 276-293.

7. Materka A. Texture analysis methodologies for magnetic resonance imaging. Dialogues Clin Neurosci 2004; 6: 243-250.

8. Castellano G, Bonilha L, Li LM, et al. Texture analysis of medical images. Clin Radiol 2004; 59: 1061-1069.

9. Tandon S, Shahab R, Benton JI, et al. Fine-needle aspiration cytology in a regional head and neck cancer center: Comparison with a systematic review and meta-analysis. Head Neck 2008; 30: 1246-1252.

10. Chmielewski A, Dufort P, Scaranelo AM. A Computerized System to Assess Axillary Lymph Node Malignancy from Sonographic Images. Ultrasound Med Biol 2015; 41: 2690-2699.

11. Jeong JJ, Lee YS, Lee SC, et al. A scoring system for prediction of lateral neck node metastasis from papillary thyroid cancer. J Korean Med Sci 2011; 26: 996-1000.

12. Desmots F, Fakhry N, Mancini J, et al. Shear wave elastography in head and neck lymph node assessment: image quality and diagnos-

tic impact compared with B-mode and Doppler ultrasonography. Ultrasound Med Biol 2016; 42: 387-398.

13. Chen SJ, Lin CH, Chang CY, et al. Characterizing the major sonographic textural difference between metastatic and common benign lymph nodes using support vector machine with histopathologic correlation. Clin Imaging 2012; 36: 353-359 e2

14. Zhang J, Wang Y, Dong Y, et al. Computer-aided diagnosis of cervical lymph nodes on ultrasonography. Computers in Biology and Medicine 2008; 38: 234-243.

15. Chang CY, Lai CC, Lai CT, et al. Integrating PSONN and Boltzmann function for feature selection and classification of lymph nodes in ultrasound images. Journal of Visual Communication and Image Representation 2013; 24: 23-30.

16. Dudau C, Hameed S, Gibson D, et al. Can contrast-enhanced ultrasound distinguish malignant from reactive lymph nodes in patients with head and neck cancers? Ultrasound Med Biol 2014; 40: 747-754.

17. Tan S, Miao LY, Cui LG, et al. Value of Shear Wave Elastography Versus Contrast-Enhanced Sonography for Differentiating Benign and Malignant Superficial Lymphadenopathy Unexplained by Conventional Sonography. J Ultrasound Med 2017; 36: 189-199.

18. Azizi G, Keller JM, Mayo ML, et al. Shear Wave Elastography and Cervical Lymph Nodes: Predicting Malignancy. Ultrasound Med Biol 2016; 42: 1273-1281.

19. Park AY, Kim JA, Son EJ, et al. Shear-Wave Elastography for Papillary Thyroid Carcinoma can Improve Prediction of Cervical Lymph Node Metastasis. Ann Surg Oncol 2016; 23: 722-729.

20. Cheng KL, Choi YJ, Shim WH, et al. Virtual touch tissue imaging quantification shear wave elastography: prospective assessment of cervical lymph nodes. Ultrasound Med Biol 2016; 42: 378-386.

21. Schürmann J. Pattern classification: a unified view of statistical and neural approaches. John Wiley & Sons, Inc., 1996.

22. Mucciardi AN, Gose EE. A comparison of seven techniques for choosing subsets of pattern recognition properties. IEEE Transactions on Computers 1971; 100: 1023-1031.

23. Van Erkel AR, Pattynama PMT. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. Eur J Radiol 1998; 27: 88-94.

24. Ophir J, Cespedes I, Ponnekanti H, et al. Elastography: a quantitative method for imaging the elasticity of biological tissues. Ultrasonic Imaging 1991; 13: 111-134.

25. Acu L, Oktar SÖ, Acu R, et al. Value of Ultrasound Elastography in the Differential Diagnosis of Cervical Lymph Nodes. J Ultrasound Med 2016; 35: 2491-2499.

26. Rubaltelli L, Beltrame V, Tregnaghi A, et al. Contrast-enhanced ultrasound for characterizing lymph nodes with focal cortical thickening in patients with cutaneous melanoma. Am J Roentgenol 2011; 196: W8-W12.

27. Lee JH, Shin HJ, Yoon JH, et al. Predicting lymph node metastasis in patients with papillary thyroid carcinoma by vascular index on power Doppler ultrasound. Head Neck 2016

28. Ardakani AA, Gharbali A, Mohammadi A. Classification of Benign and Malignant Thyroid Nodules Using Wavelet Texture Analysis of Sonograms. J Ultrasound Med 2015; 34: 1983-1989.

29. Zhang J, Wang Y, Yu B, et al. Application of Computer-Aided Diagnosis to the Sonographic Evaluation of Cervical Lymph Nodes. Ultrasonic Imaging 2016; 38: 159-171.