Pattern Recognition 43 (2010) 3494–3506

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: [www.elsevier.com/locate/pr](http://www.elsevier.com/locate/pr)

Application of support-vector-machine-based method for feature selection and classiﬁcation of thyroid nodules in ultrasound images

Chuan-Yu Chang a, Shao-Jer Chen b,c,×, Ming-Fong Tsai a

a *Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliou, Yunlin, Taiwan*

b *Department of Medical Imaging, Buddhist Dalin Tzu Chi General Hospital, Chia-Yi, Taiwan*

c *School of Medicine, Buddhist Tzu Chi University, Hualien, Taiwan*

a r t i c l e i n f o

*Article history:*

Received 24 September 2009 Received in revised form

27 March 2010

Accepted 30 April 2010

*Keywords:*

Support vector machines Feature selection

Thyroid nodule classiﬁcation

a b s t r a c t

Most thyroid nodules are heterogeneous with various internal components, which confuse many radiologists and physicians with their various echo patterns in ultrasound images. Numerous textural feature extraction methods are used to characterize these patterns to reduce the misdiagnosis rate. Thyroid nodules can be classiﬁed using the corresponding textural features. In this paper, six support vector machines (SVMs) are adopted to select signiﬁcant textural features and to classify the nodular lesions of a thyroid. Experiment results show that the proposed method can correctly and efﬁciently classify thyroid nodules. A comparison with existing methods shows that the feature-selection capability of the proposed method is similar to that of the sequential-ﬂoating-forward-selection (SFFS) method, while the execution time is about 3–37 times faster. In addition, the proposed criterion function achieves higher accuracy than those of the *F*-score, *T*-test, entropy, and Bhattacharyya distance methods.

& 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Nodular lesions of the thyroid are very common in the general population. The incidence of palpable thyroid nodules in the adult population is about 4–8%. The incidence of thyroid nodules detected by autopsy is much higher (50%) [[1]](#_bookmark16). Compared with magnetic resonance imaging (MRI) and computerized tomography (CT), sonography is chosen for the diagnosis of thyroid nodules due to its sensitivity and convenience [[1,2]](#_bookmark16). Most thyroid nodules tend to have various internal echogenicities in the sonogram, which makes a deﬁnite diagnosis difﬁcult. If the characteristic echogeni- cities for the major components of the thyroid nodule can be identiﬁed, the interpretation of thyroid sonography can be more accurate, which would decrease the misdiagnosis rate of thyroid cancer.

Sonographic ﬁndings of nodular lesions, such as nodular goiter and thyroid tumors, are well documented in textbooks and many articles [[3–6]](#_bookmark17). In nodular goiter, the sonographic changes are usually heterogeneous. In follicular adenoma, changes are either isoechoic or hyperechoic. Several ultrasound features have been found to be associated with an increasing risk of thyroid cancer, including hypoechogenicity and predominantly solid composition; however, no ultrasound feature has both a high sensitivity and a high positive predictive value for thyroid cancer [[1]](#_bookmark16).

× Corresponding author. Tel.: + 866 5 2648000x5750; fax: + 886 5 2648555.

*E-mail address:* [shaojer.chen@msa.hinet.net (S.-J. Chen).](mailto:shaojer.chen@msa.hinet.net)

Fine-needle aspiration (FNA) is a critical diagnostic test for determining the proper management of thyroid nodular disease following thyroid ultrasonography. Except for almost entirely cystic change, nodules larger than 1 cm should be evaluated because they have the potential to be clinically signiﬁcant cancers [[1,2]](#_bookmark16). However, many physicians are confused about the nature of various echo patterns of thyroid nodules. Approximately, 10–20% of thyroid biopsies by FNA are nondiagnostic [[7]](#_bookmark18). Hence, in this paper, a helpful thyroid nodule classiﬁcation system that properly classiﬁes the thyroid nodules in thyroid ultrasonography and assists physicians in ascertaining thyroid nodules before FNA.

Various feature extraction methods have been recently proposed [[8–17]](#_bookmark19). A lot of features can thus be obtained from medical images. However, extracting these features and using them to directly train a classiﬁer is very time-consuming. To reduce the time required and to improve accuracy, it is necessary to reduce the number of features; however, it is difﬁcult to select the most signiﬁcant ones. Classiﬁcation and prediction using signiﬁcant features can obtain higher accuracy than that obtained with all features. Therefore, a powerful classiﬁer and an efﬁcient feature selection method are required. Support vector machines (SVMs) proposed by Vapnik et al. are supervised learning machines [[18]](#_bookmark27). Since SVM is capable of generating a hyperplane to separate two data sets and of providing good generalization, it is a powerful classiﬁcation approach. Recently, Tsantis et al. proposed an SVM-based method for assessing the malignancy risk of thyroid nodules [[27]](#_bookmark31) with a maximum classiﬁcation accuracy of 96.7%. Wang et al. adopted SVM to establish a diagnostic model [[28]](#_bookmark33).

0031-3203/$ - see front matter & 2010 Elsevier Ltd. All rights reserved. doi:10.1016/j.patcog.2010.04.023

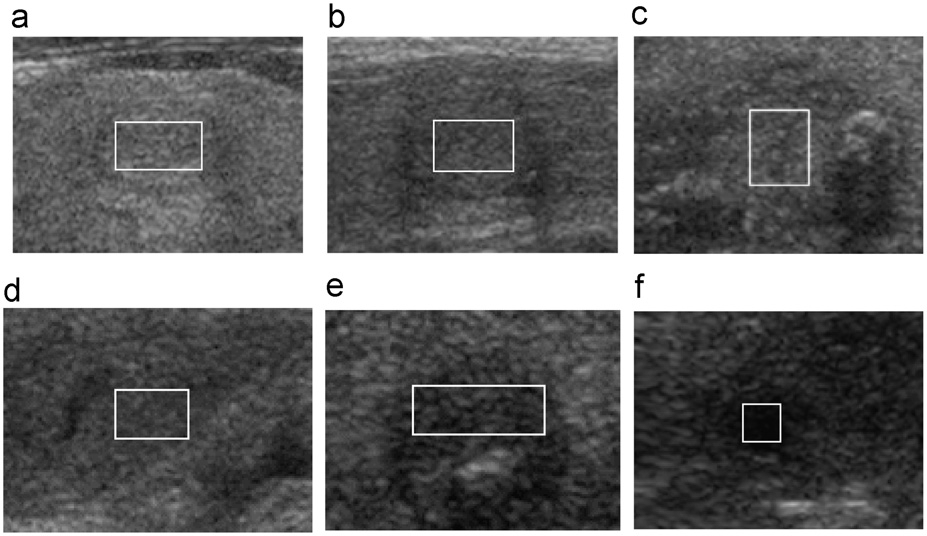


Fig. 1. Six ROIs of thyroid nodules were outlined by an experienced radiologist and conﬁrmed by biopsies: (a) enlarged follicles, (b) follicular cells with follicles,

(c) papillary cells with follicles, (d) follicular cells with ﬁbrosis, (e) papillary cells with ﬁbrosis, and (f) ﬁbrosis.

The diagnostic model combined with 3 biomarkers could differentiate thyroid cancer from thyroid adenoma with a speciﬁcity of 88.9% and a sensitivity of 96.9%. In the present

study, SVMs are applied to select signiﬁcant features and to

response. SVM solves the following primal problem [[26]](#_bookmark33):

*l*

1 X

min *wT w* þ*C* x

*w*,*b*,x

2

*i*

*i* ¼ 1

classify the thyroid nodules into six categories.

[Fig. 1](#_bookmark0) shows six types of ultrasound image of thyroid nodules with the region of interest (ROI), which has been outlined by a radiologist and conﬁrmed by biopsy. To recognize thyroid nodules, 78 textural features are extracted from each ROI. SVMs and the proposed feature selection method are then applied to select the signiﬁcant features. Each SVM is trained by speciﬁc features, which have high discrimination between two types of

subject to *yi*ð*wT* fð*xi*Þþ *b*ÞZ1—x*i*

x*i* Z 0, *i* ¼ 1,::,*l* ð1Þ

Here, training vectors *xi* are mapped into a higher dimensional space by the function f. The weight vector w and the bias *b* deﬁne the separating hyperplane. The SVM then ﬁnds a linear separating

hyperplane with the maximal margin in this higher dimensional space. The dual problem of Eq. (1) is deﬁned as follows:

thyroid nodule. Finally, the trained SVMs are used to classify the

min

a

1a*T Q* a—*eT* a

ROI of the thyroid nodule.

2

The rest of this paper is organized as follows. Section 2 describes related approaches of feature selection, which includes support vector machines, feature evaluation, and feature extrac- tion. In Section 3, the modiﬁcation of feature evaluation and a combination of support vector machines is proposed. The SVM-based method for feature selection is implemented to classify

thyroid nodules in ultrasound images. Section 4 presents the

subject to *yT* a ¼ 0

0 rar *C*, *i* ¼ 1, ... ,*l* ð2Þ

where *e* is a vector of all ones, *C* is the penalty parameter of the error term, *C*40 is the upper bound of all variables a, and *Q* is an *l*-by-*l* positive semideﬁnite matrix. Note that *Qij*÷*yiyjK*(*xi*, *xj*), where *K*(*xi*,

*x* ) f(*x* )*T*f(*x* ), is called the kernel function. Then *w* P*l* a *y* f *x*

experiment results obtained from the proposed method and a

comparison with other feature selection methods. Finally, conclu-

*j* ÷

*i*

*j*

¼

*i i*

ð *i*Þ

*i* ¼ 1

and *sgn wT* f *x b sgn*. P*l* a *y K x* ,*x b*Σ is the decision func-

sions are given in Section 5.

ð

ð Þþ Þ¼

*i* ¼ 1

*i i*

ð *i*

Þþ

1. Support vector machines, feature extraction, and feature

selection

This section describes the pre-processing and related works, including support vector machines, feature extraction, and feature evaluation. Methods for extracting spatial and frequency features from an image are presented. Since there is a large number of extracted features, feature evaluation is used to reduce the number of features which are used to train SVMs and to retain signiﬁcant features for the subsequent classiﬁcation of thyroid nodules.

* 1. *Support vector machine*

tion. In this paper, SVMs are implemented using LIBSVM [[24]](#_bookmark31), and the

kernel function is the radial basis function (RBF) deﬁned as follows:

*K*ðxð*s*Þ,x Þ¼ exp.—:xð*s*Þ—x :2=2s2 Σ ð3Þ

*i*

*i*

* 1. *Feature evaluation based on statistical criterion*

78 textural features are obtained in the feature extraction stage. Although they can be directly used to classify thyroid nodules, the time required for classiﬁer training increases with the number of features. When the range of ROI is wide, the time required for calculating the features is long. Hence, this section introduces some methods for feature selection which reduce redundant and insigniﬁcant features.

Feature selection is one of the main problems in machine learning and statistics. As described above, SVM is an effective

Consider the training set fðx*i*,*yi*Þg*N*

, where *N* is the number of

classiﬁer in machine learning, but it cannot be used to directly

training samples, *xi*A*Rn*

*i* ¼ 1

is the *n*-dimensional input space for the

obtain the importance of a feature. Hence, some statistical criteria

*i*th sample, and *yi*A{+1, — 1} is the corresponding desired

are introduced to rank or score each feature. The *F*-score, *T*-test,

entropy, and the Bhattacharyya distance of the *i*th feature are calculated, respectively, as [[19,21]](#_bookmark28) follows:

* 1. *Feature extraction*

*F*ð*i*Þ¼

.*x*ðþ Þ *x* Σ2

.*x*ð—Þ *x* Σ2

ð4Þ

In the proposed method, 78 textural features are extracted from each ROI of the thyroid ultrasound images. The ROI is

.sðþ ÞΣ2 þ.sð—ÞΣ2

i — *i* þ

i — *i*

*i i*

*T*ð*i*Þ¼ .rﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃ*x*ð*i* þ Þ —ﬃﬃ *x*ð*i*—Þ .

.

=*n*—.

. . Σ2ﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃ . Σ2 .

# 

sð*i* þÞ

=*n* þ þ

sð*i*—Þ

ð5Þ

outlined by an experienced radiologist and conﬁrmed by biopsy. The extracted textural features are as follows:

1. *Gray level co-occurrence matrix*: The gray level co-occurrence

matrix (GLCM), which corresponds to second order statistics,

was proposed by Haralick et al. [[8]](#_bookmark19). It is a two dimensional

*E*ð*i*Þ¼ 1Σ.sðþÞΣ2 =.sð—ÞΣ2 þ.sð—ÞΣ2 =.sðþ ÞΣ2 —2

2

*i*

*i*

*i*

*i*

array in which both rows and columns represent a set of

possible image values [[9]](#_bookmark20). There are two important para-

.*x*ðþÞ *x*ð—ÞΣ2 .1=.sðþ ÞΣ2

1=.sð—ÞΣ2ΣΣ

6 meters for constructing GLCM: distance *d* and angle *a*. There

þ *i* — *i*

*i* þ *i*

ð Þ are two pixels with distance d between them in the direction

speciﬁed by the angle a. For example, *I*(*x*, *y*) denotes a pixel at

.

Σ2

1

2 . Σ2 . Σ2 3

position (*x*, *y*) in an image *I*. The distance between *I*(*x*, *y*) and

*x*ð*i* þ Þ—*x*ð*i*—Þ

sð*i*—Þ

# 

4

sð*i* þ Þ

þ

sð*i* þ Þ

6

sð*i* þ Þ

2

þ sð*i*—Þ

7

sð*i*—Þ

its 8 adjacent pixels is equal to 1. According to this distance,

*B*ð*i*Þ¼ Σ.

Σ2 .

Σ2 Σ þ 2 ln4 rﬃ.ﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃΣﬃﬃ2ﬃﬃﬃ.

Σ2 5 ð7Þ

*I*(*x*, *y*), *I*(*x* +1, *y*), and *I*(*x* — 1, *y*) have a 01 angular relationship.

where *xi*, *x*ð*i* þ Þ , and *x*ð*i*—Þ are the averages of the *i*th feature of the

*I*(*x*, *y*), *I*(*x* +1, *y* — 1), and *I*(*x* — 1, *y* — 1) have a 451 angular

relationship. 901 and 1351 relationships are also possible.

whole, positive, and negative instances, respectively. *n* + and *n*— are the numbers of positive and negative instances, respectively. sð*i* þ Þ and sð*i*—Þ are the standard deviations of the *i*th feature of the

positive and negative instances, respectively.

The feature vectors must be normalized before they are evaluated using these criteria. After ranking, all feature vectors are used to train the SVM. Then, a *k*-fold cross-validation method is applied to evaluate the performance of each combination of features, and a grid search algorithm is applied to ﬁnd the best

parameters *C* and g to improve the classiﬁed accuracy of the SVM

[[19]](#_bookmark28). *C* is a penalty parameter which was introduced in Section 2, andg denotes the term 1/2s2 of the RBF kernel function. *C* and g

are selected using a grid search algorithm; the ones with the best validation rate are chosen. The procedure is as follows:

1. Consider the grid space of (*C*,g) with log2*C*A{ — 2, 0, y, 12} and log2gA{ — 2, 0, y, 12}.
2. For each hyperparameter pair (*C*,g) in the search space,

conduct the *k*-fold cross-validation on the training set.

1. Choose the parameters (*C*,g) that leads to the lowest error validation rate.
2. Use the best parameters to create a model as the classiﬁer.

After each process of cross-validation, the validation rate is recorded and the feature with the lowest score is eliminated. The remaining features are used to train the SVM and to perform cross-validation again. These procedures are repeated until all features are eliminated. Finally, the combination with the highest validation rate and with the least features is selected. The procedure is as follows [[19,20]](#_bookmark28):

1. Evaluate and sort each feature using the statistical criterion.
2. Drop features below some thresholds, and then:
   1. Randomly split the training data into *Xtrain* and *Xvalid*.
   2. Let *Xtrain* be the new training data, and perform the *k*-fold cross-validation to predict *Xvalid* using SVM.
   3. Repeat the above steps to calculate the average validation rate until there are no features to drop.
3. Choose the best feature subset with the largest average validation rate.
4. Retrain the SVM with the best feature subset, and predict the accuracy with the test set.

A 4-fold cross-validation is adopted in this paper.

Accordingly, 13 features are extracted from GLCM [[8,10]](#_bookmark19). They

are as follows: (F1) correlation, (F2) difference entropy, (F3) difference variance, (F4) sum average, (F5) sum entropy, (F6) sum of squares, (F7) sum variance, (F8) contrast, (F9) energy, (F10) entropy, (F11) local homogeneity, (F12) cluster shade, and (F13) cluster prominence.

By using various values of *d* and *a*, many GLCMs can be constructed to obtain a lot of features. To eliminate the attenuation effect during the transmission and reﬂection of the ultrasound, the angle is set to 01 in this study. Moreover, in order to preserve the complexity of the spatial relationship, a displacement vector of distance ¼ 1 is set.

1. *Statistical feature matrix*: The statistical feature matrix (SFM) was proposed by Wu and Chen [[11]](#_bookmark21). It is constructed using operations between pixels in an image. The (F14) dissimilarity extracted from SFM is used for statistical analysis [[11,12]](#_bookmark21).
2. *Gray level run-length matrix:* The gray level run length matrix (GLRLM) was proposed by Galloway [[13]](#_bookmark22). Similar to GLCM, GLRLM is constructed using distance and angle. The distance represents how long a speciﬁc gray level can appear continuously from one pixel to another in a user-speciﬁed direction with angle *a*. Let *C* denote the longest distance that a speciﬁc gray level appears, and *H* and *W* be the height and width of the ROI, respectively. Angular parameter *a*¼ 01and *C H* þ *W* c are used in this study.

¼b

pﬃﬃﬃﬃ2ﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃﬃ2ﬃﬃ

Five features, (F15) short runs emphasis, (F16) long runs emphasis, (F17) gray-level uniformity, (F18) run length uniformity, and (F19) run percentage are calculated from

the gray level run-length matrix [[14]](#_bookmark23).

1. *Laws’ texture energy measures*: Laws developed a texture-energy approach that measures the amount of variation within a ﬁxed- size window [[9]](#_bookmark20). This approach has three simple vectors: *L*3¼ (1, 2, 1), *E*3¼ (— 1, 0, 1), and *S*3¼ (— 1, 2, — 1). The *L*3 vector represents a center-weighted local averaging. The *E*3 vector detects the edges and the *S*3 vector detects the spots. The vectors are convoluted with themselves or with each other. Then, three new vectors are obtained: *L*5¼ *L*3\**L*3¼ (1, 4, 6, 4, 1), *L*5¼ *L*3\**E*3¼ (— 1, — 2, 0, 2, 1), and *S*5¼ *L*3\**S*3¼ (— 1, 0, 2, 0, — 1), where \* is the convolution operator. After obtaining the three new vectors, ﬁve Laws’ 5 ~ 5 masks can be generated: *LE*¼ *L*5 *E*5, *EL*¼ *E*5 *L*5, *SL*¼ *S*5 *L*5, *EE*¼ *E*5 *E*5, and *LS*¼ *L*5 *S*5. Finally, the masks are convoluted with the image and the statistics from the results are used.

T T T T T

The features calculated from Laws’ texture energy measures are (F20) *LE* mean, (F21) EL mean, (F22) *SL* mean, (F23) EE mean,

(F24) LS mean, (F25) LE variance, (F26) EL variance, (F27) SL variance, (F28) EE variance, and (F29) *LS* variance.

1. *Neighboring gray level dependence matrix*: The neighboring gray level dependence matrix (NGLDM) is a two-dimensional matrix constructed using the gray level relationship between each pixel and its neighbors in an image [[15]](#_bookmark24). Each element within the matrix is computed by a mask, which is deﬁned by the distance parameter *d*. For example, if *d* is equal to 1, a 3- by-3 mask can be obtained, if *d* is equal to 2, a 5-by-5 mask can be obtained, and so on. In this method, the relationship between pixels can be considered within different distances. For instance, the spatial relationship between the central pixel and pixels within a distance equal to 1 or 2 can be determined by using a 5-by-5 mask. Furthermore, each difference of gray level which is equal to or less than a user-speciﬁed factor *a* between the center pixel and its neighbors is computed using a mask in the image.

The following features are calculated from this matrix: (F30) small number emphasis, (F31) large number emphasis, (F32)

number nonuniformity, (F33) second moment, and (F34) entropy. Many combinations of the two parameters can be used to obtain the matrix. In this study, the *d* ¼ 1 and *a*¼ 0.

1. *Wavelet features*: An image can be decomposed into low and high frequency sub-bands using one-dimensional lowpass and highpass ﬁlters, respectively. Another decomposition produces four sub-bands, *LL*, *LH*, *HL*, and *HH*. The coefﬁcients of the lowpass and highpass ﬁlters were deﬁned by Antonini et al. [[16]](#_bookmark25). Because the *LL* sub-band is an approximation of the image, it is used to calculate the textural features.

The mean, standard deviation, and Laws’ features of the *LL* sub-band image are calculated as wavelet features. The extracted features are (F35) mean, (F36) standard deviation, (F37) *LL LE* mean, (F38) *LL EL* mean, (F39) *LL SL* mean, F40) *LL*

*EE* mean, (F41) *LL LS* mean, (F42) *LL LE* variance, (F43) *LL EL*

variance, (F44) *LL SL* variance, (F45) *LL EE* variance, and (F46)

*LL LS* variance.

1. *Fourier feature based on local Fourier coefﬁcients*: The Fourier transform is applied to obtain the local Fourier coefﬁcient maps of the image [[17]](#_bookmark26). Each coefﬁcient consists of two parameters: magnitude and phase angle. The histograms of the two parameters can thus be obtained. Finally, the mean and standard deviation of the magnitude and phase angle of local Fourier coefﬁcient maps are computed as textural features. (F47–F54) are the means of 8 magnitudes, (F55–

F62) are the means of 8 phase angles, (F63–F70) are the

standard deviations of 8 magnitudes, and (F71–F78) are the standard deviations of 8 phase angles.

1. Proposed method

This section describes the proposed method, including a ﬂowchart of the proposed system, the *k*-fold feature selection method, and the SVM-based thyroid nodule classiﬁcation. The *k*-fold feature selection method uses a concept similar to that of *k*-fold cross-validation; it is combined with feature evaluation for

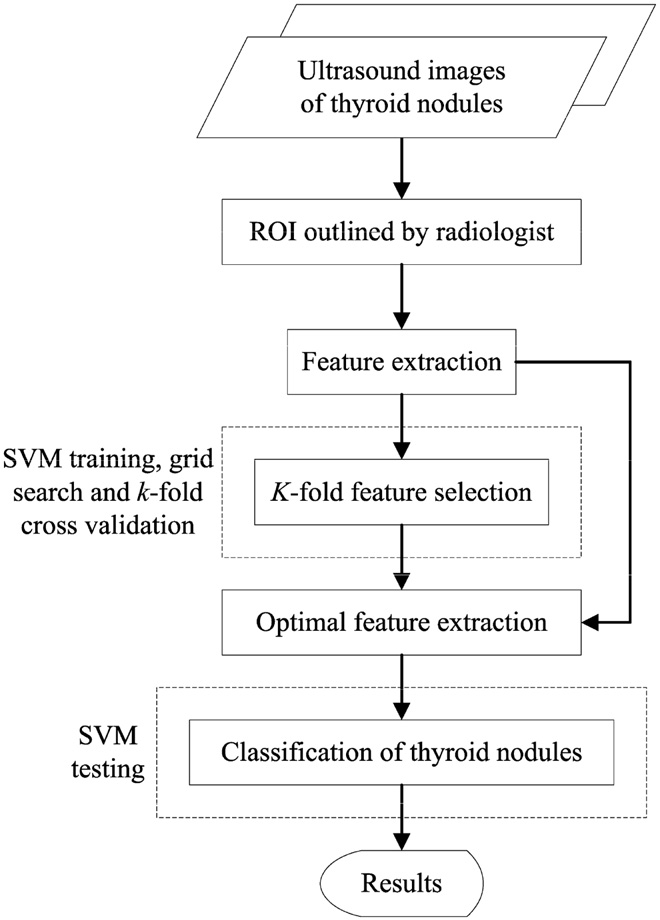


Fig. 2. Flowchart of the proposed system.

In the selection stage, the feature vectors are used to train the SVM and to perform *k*-fold cross-validation to evaluate the performance of each feature subset. Once the best feature subset is obtained, the optimal features can be extracted from the original feature vectors accordingly. These optimal features are used to retrain the SVM for the classiﬁcation of thyroid nodules.

* 1. *k-Fold feature selection*

The concept of *k*-fold feature selection is similar to that of the *k*-fold cross-validation. The data set is separated into *k* folds, in which the ﬁrst fold is regarded as the testing data, and the remaining folds are the training data. The training data are used to train the SVM. [Fig. 3](#_bookmark2) shows a ﬂowchart of *k*-fold feature selection. As shown in the ﬁgure, the testing set is not used until the best feature set is obtained; i.e., it is used to perform the classiﬁcation of thyroid nodules.

Once the training data are ready, the signiﬁcance of features is evaluated by *F*-score (Eq. (4)), *T*-test (Eq. (5)), and the following:

j*x*ðþ Þ—*x*ð—Þj

selecting the discriminative feature set. Then, the SVM-based Correlation : *P*ð*i*Þ¼ *i i*

ð8Þ

method classiﬁes thyroid nodules using the selected features.

*i*

*i*

.sðþ ÞΣ2 .sð—ÞΣ2

*3.1. System architecture*

[Fig. 2](#_bookmark1) shows a ﬂowchart of the proposed system. The sonographic thyroid nodular images are collected and outlined by a radiologist and conﬁrmed by biopsies. Then, the features of ROIs are extracted and combined as feature vectors. These feature vectors are separated into *k* folds for feature selection.

where *x*ð*i* þ Þ and *x*ð*i*—Þ are the averages of the *i*th feature of the positive and the negative instances, respectively. sð*i* þ Þ and sð*i*—Þ are the standard deviations of the *i*th feature of the positive and the

negative instances, respectively. Then, the feature scores are ranked individually in descending order. Three ranking lists are thus obtained: the *F*-score ranking list, the *T*-test ranking list, and the correlation ranking list using Eqs. (4), (5), and (8), respec- tively. Then, the intersectional features of the top *N*% of features

Testing set

Training set

Calculate the scores of the features using Eq. (4),

(5) and (8), and sort them in descending order

The *k* folds are used as the training and testing sets alternatively.

Sort the intersectional features using Eq.

(9) in ascending order.

Keep the top *N*% features and find the intersectional features of the ranking lists from Eq. (4) and (5).

Find the intersectional features of the ranking lists from Eq. (8) and (9).

Sort the intersectional features in the same order of the ranking list which was obtained from Eq. (9).

Training set with feature subset *f.*

Validation rate

*f* > 0 ?

Yes

No

Best feature subset

*f* = *f* -1

Using *k*-fold cross validation to train and test, and perform grid search algorithm to find the best parameters.

Generate new data set with the number of features *f. f* is the total number of intersectional features.

Fig. 3. *k*-Fold feature selection ﬂowchart.

are kept. The intersectional features are the features that are present in both the *F*-score ranking list and the *T*-test ranking list. [Fig. 4](#_bookmark3) shows an example of ﬁnding the ﬁnal ranking list of features. [Fig. 4](#_bookmark3)(a) shows the top 10 features of the *F*-score ranking list, the *T*-test ranking list, and the correlation ranking list. Firstly, the intersectional features of the *F*-score ranking list and the *T*-test ranking list are kept using *F*-score ranking list-major and *T*-test ranking list-major, respectively. The two intersectional feature lists are shown in [Fig. 4](#_bookmark3)(b) are thus obtained. Then, the following equation is used to re-weight these intersectional features, which are then sorted in ascending order to make up



ROIs of thyroid nodules

1st fold

2nd fold

*K*th fold

the weighted ranking list:

*W*ð*i*Þ¼ *wF FP*ð*i*Þþ *wT TP*ð*i*Þ ð9Þ

where *FP*(*i*) and *TP*(*i*) are the places of the *i*th intersectional feature in the ranking lists which were obtained from *F*-score (Eq. (4)) and *T*-test (Eq. (5)), respectively. *wF* and *wT* are adjustable weights. According to experiments, the recommended weight range for *wF* and *wT* is [[1,3]](#_bookmark16) for a reduced search time. The weighted ranking list is used to ﬁnd the intersectional features with the correlation ranking list. These features are sorted in the same order as that of the weighted ranking list to generate the ﬁnal ranking list; i.e., the ﬁnal ranking list is obtained by extracting the features that are present in both the weighted list and the correlation ranking list using weighted list-major.

Once the ﬁnal ranking list is obtained, a new training data set is obtained with the resulting number of features to train the SVM. The *k*-fold cross-validation and grid search are performed, and the validation rates are recorded. In each validation, the last- place feature is eliminated. The remaining features are used to train the SVM and to perform cross-validation and grid search. These procedures are repeated until all features are eliminated. Finally, the best feature subset, which satisﬁes the following conditions, is kept:

1. It has the highest validation rate.
2. It has the lowest number of features.

4-Fold feature selection is adopted in this paper.

* 1. *Classiﬁcation of thyroid nodules*

Six binary-SVMs [[24]](#_bookmark31) are used to classify thyroid nodules into seven categories. [Fig. 5](#_bookmark4) shows the structure of the six binary-SVMs used for thyroid nodule classiﬁcation. The most discriminative feature combinations are used to train the classiﬁers.

The ﬁrst binary-SVM (S0) is used to categorize the ROIs as thyroid nodules or non-nodules. According to pathology, the thyroid nodules can be categorized as follicles-based or ﬁbrosis- based. The second binary-SVM (S1) is used to classify the nodules.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | F-score Ranking List |  |  |  |  |  | Intersectional features in F-score ranking list |
| Rank | 0 1 2 3 4 5 | 6 | 7 | 8 | 9 |  | Rank 0 1 2 3 4 5 6 7 |
| No. of features | 35 47 4 10 7 12 | 69 | 65 | 2 | 9 |  | No. of features 35 47 4 10 7 12 69 65 |
|  | T-test Ranking List |  |  |  |  |  | Intersectional features in T-test ranking list |
| Rank | 0 1 2 3 4 5 | 6 | 7 | 8 | 9 |  | Rank 0 1 2 3 4 5 6 7 |
| No. of features | 35 7 47 4 12 10 | 6 | 13 | 69 | 65 |  | No. of features 35 7 47 4 12 10 69 65 |
|  | Correlation Ranking List |  |  |  |  |  |  |

Rank 0 1 2 3 4 5 6 7 8 9

No. of features 9 13 6 12 59 10 65 69 53 49



*W* (47)  11  2  2  5,*W* (7)  1 4  2 1  6



Correlation Ranking List

Rank 0 1 2 3 4 5 6 7 8 9

Weighted Ranking List

Rank 0 1 2 3 4 5 6 7

No. of features 35 47 7 4 10 12 69 65

# 

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. of features | 9 | 13 | 6 | 12 | 59 | 10 | 65 | 69 | 53 | 49 |

Weighted Ranking List

Rank 0 1 2 3 4 5 6 7

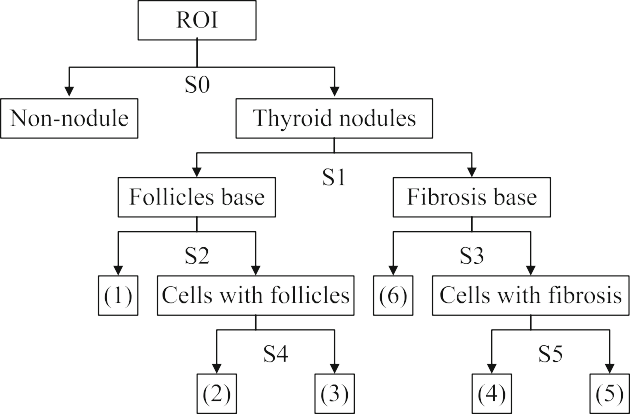
Final Ranking List

Rank 0 1 2 3

No. of features 10 12 69 65

No. of features 35 47 7 4 10 12 69 65

Fig. 4. An example of ranking and selecting intersectional features: (a) three top 10 features in *F*-score, *T*-test, and correlation ranking lists. (b) Intersectional features of *F*-score and *T*-test ranking lists. (c) features are calculated using Eq. (8) with *wT* ¼ 1 and *wF* ¼ 2 in (b). (d) According to the results in (c), the intersectional features are sorted again in an ascending order to generate the weighted ranking list. (e) The intersectional features of the correlation ranking list and the weighted ranking list are marked in gray. (f) Intersectional features are sorted again in the order of the weighted ranking list to obtain the ﬁnal ranking list.

* 1. *Accuracy measurement*

To evaluate the performance of the classiﬁcation, the valida- tion accuracy and the testing accuracy are, respectively, deﬁned as follows:

*k*

P

Validation accuracy ¼ *i* ¼ 1

*CVTPi* þ

*CVN*

*CVTNi*

ð10Þ

Testing accuracy *TP* þ *TN*

¼

*N*

ð11Þ

Fig. 5. Structure of the six binary-SVMs used for thyroid nodule classiﬁcation. Nodes (1)–(6) are the corresponding nodules of [Figs. 1](#_bookmark0)(a)–(f), respectively.

Follicles base nodules can be further categorized as enlarged follicles and cells with follicles. The third binary-SVM (S2) is used to classify these nodules. Cells with follicles can be further categorized as follicular cells with follicles and papillary cells with follicles, which are classiﬁed by S4.

Fibrosis base nodules can be categorized as cells with ﬁbrosis and the pure ﬁbrosis. The fourth binary-SVM (S3) is used to classify these nodules. Cells with ﬁbrosis can be further categorized as follicular cells with ﬁbrosis and papillary cells with ﬁbrosis. These cells are classiﬁed using the last SVM, S5. Accordingly, the thyroid nodules can be separated into seven categories using the six binary-SVMs.

where *CVTPi* and *CVTNi* are the numbers of true positive and true negative instances of the *i*th fold of the *k*-fold cross-validation, respectively. *TP* and *TN* are the numbers of true positive and true negative instances of the testing pattern, respectively. *CVN* and *N* are the total numbers of the validation and testing patterns, respectively.

1. Experiment results

From January 2005 to March 2007, 61 patients (48 females and 13 males, age range: 23–82 years old) with 76 thyroid nodular lesions were studied. The number of nodules, their pathology classiﬁcation, patient source (indicated by the number of cases), and ﬁnal diagnosis are shown in [Table 1](#_bookmark5). There is more than one nodule in some cases. The range of the number of nodules detected in each case is one to three. Sonographic images of the nodular lesions were obtained before operation with a commercial sonographic system (GE LOGIQ 700 ultrasound system, General Electric Healthcare, Chalfant St. Giles, UK; System 1). The probe

was a B-mode linear array with an operation frequency range of 5–10 MHz. Parameters of image acquisition were kept constant, including the time-gain compensation and focal zone (7 focal zones within a 2-cm depth range). Other parameters were set as follows: dynamic range, 78 dB; gain, 34; edge enhance, E2; gray map, MC; frame average settings, A2 [[23]](#_bookmark30). The image which had the maximum longitudinal section of each thyroid nodule was selected.

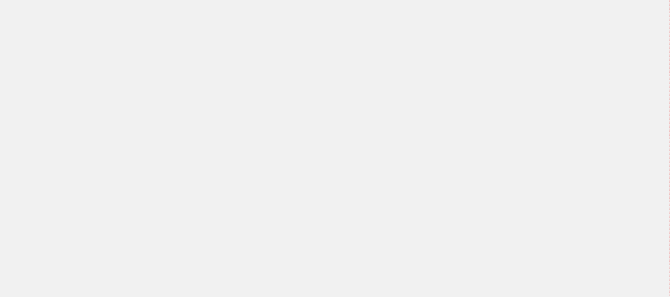
To evaluate the generalization of the proposed method, sonographic images obtained from another sonographic system (system 2) were also used for testing. A reasonable range of image acquisition settings is tolerated. In system 2, a B-mode linear array probe (i12L) with an operation frequency of 6–11 MHz was used for evaluation [[23]](#_bookmark30). Other parameters were set as follows: dynamic range, 69 dB; gain, 34; edge enhance, E3; gray map, MC; frame average settings, A3 [[23]](#_bookmark30).

To select the best feature set for systems 1 and 2, a total of 631 and 1008 ROIs obtained from systems 1 and 2, respectively, were used to train and test the SVMs. They included 324 and 540 ROIs of non-nodules, 59 and 73 ROIs of enlarged follicles, 82 and 95 ROIs of follicular cells with follicles, 59 and 78 ROIs of papillary cells with follicles, 27 and 101 ROIs of follicular cells with ﬁbrosis, 30 and 87 of papillary cells with ﬁbrosis, and 50 and 34 ROIs of ﬁbrosis obtained from systems 1 and 2, respectively. In addition, to ﬁnd a common discriminative feature subset between systems 1 and 2, the feature vectors obtained from the two systems were mixed to train and test the SVMs [[24,26]](#_bookmark31). The testing accuracies and computational times with the most discriminative feature subsets (extracted from systems 1, 2, and 1 + system 2) of each SVM are shown in [Table 2](#_bookmark6).

From [Table 2](#_bookmark6), the proposed method can extract the most discriminative feature (F14) for S0 to classify nodular and

Table 1

Number of nodules, their pathology classiﬁcation, patient source (indicated by the number of cases), and ﬁnal diagnosis.



|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
| Pathology classiﬁcation | Number nodules | of | Number of cases/diagnosis |
| Enlarged follicles | 13 |  | 12/nodular goiter |
| Follicular cells with follicles | 22 | 15/nodular goiter, |
|  |  | 5/adenoma |
| Papillary cancer cells with | 10 | 8/papillary cancer |
| follicles |  |  |
| Follicular cells with ﬁbrosis | 10 | 5/nodular goiter |
| Papillary cancer cells with | 7 | 5/papillary cancer |
| ﬁbrosis |  |  |
| Fibrosis | 14 | 7/nodular goiter, 4/papillary |
|  |  | cancer |
|  | | | |

non-nodular regions in system 1, system 2, and system 1 + system 2 with an accuracy of 100%. The proposed method also extracted the most discriminative features for S1 and S2 to classify follicles- based/ﬁbrosis-based, and cells with follicles in system 1, system 2, and system 1 + system 2 with an accuracy of 100%.

* 1. *Comparison of feature selection*

To evaluate the testing accuracies of the proposed method, *F*-score, *T*-test, entropy, and the Bhattacharyya distance, a *k*-fold cross-validation, were applied. The *k* value affects the number of training and testing data points. In order to obtain enough training data points while maintaining sufﬁcient testing data, *k* is set to 4 in this paper. The testing accuracies with the most discriminative feature subsets of each SVM and of each method are shown in [Table 3](#_bookmark7). The table shows that the non-nodule regions can be correctly classiﬁed and that each category of thyroid nodule can be effectively classiﬁed using the proposed method with one or two features. A comparison with other methods shows that the proposed method uses an equal or lower number of features to classify the six categories of thyroid nodule with an equal or higher accuracy. The highest validation accuracy of each validation is 100%.

From [Table 3](#_bookmark7), the (F14) dissimilarity extracted from the statistical feature matrix is explicitly a signiﬁcant feature for classifying nodular and non-nodular regions. (F14) was selected by all methods. The (F7) sum variance and (F10) entropy derived from GLCM can well reﬂect echogenicity and can effectively differentiate between follicle- and ﬁbrosis-based thyroid nodules. The (F18) run length uniformity calculated from the gray level run-length matrix is a signiﬁcant feature that is selected by the proposed, *F*-score, *T*-test, and Bhattacharyya distance criteria for classifying cells with follicles and cells with ﬁbrosis.

The number of features obtained by each method is shown in [Table 4](#_bookmark8). The proposed method selected the most concentrated features. In [Tables 3 and 4](#_bookmark7), the best features satisfy the conditions described in Section 3.

The proposed method was compared with the widely used approach proposed by Pudil et al. [[22]](#_bookmark29), called sequential ﬂoating forward selection (SFFS). The terminative condition of the SFFS algorithm is the number of features, which is determined by users. However, in practice, users usually have no idea how many features will result in a higher accuracy. Thus, the number of features was set to be equal to that of the proposed method. A comparison of the proposed method and SFFS is shown in [Table 5](#_bookmark9). The table shows that the proposed method and SFFS have the same validation and testing accuracies. However, SFFS required

Table 2

Comparison of testing accuracies and most discriminative feature subsets of each SVM and of system 1, system 2, and system 1 + system 2.

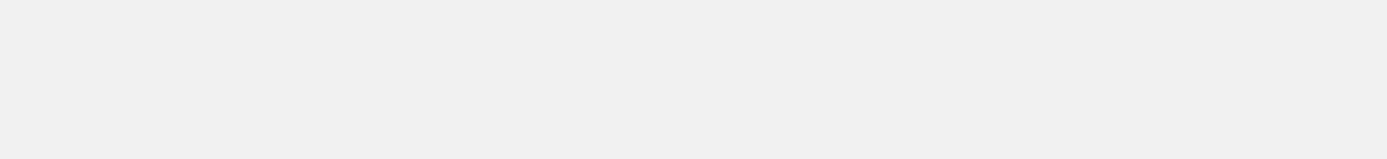
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SVM no.  S0 S1 S2 S3 S4 S5 | | | | | | | | | | | | |
| Proposed | method | (system 1) | Best features |  | F14 | F5, F7 | F1, F24 | F54, F48, | F5, | F21 | F30, F20 | F41 |
|  |  |  | Testing accuracy | (%) | 100 | 100 | 100 | 96.43 |  |  | 97.14 | 93.33 |
|  |  |  | Time (s) |  | 44.36 | 44.468 | 124.921 | 53.297 |  |  | 41.954 | 26.891 |
| Proposed | method | (system 2) | Best features |  | F14 | F10, F69 | F17, F18 | F9 |  |  | F33 | F17 |
|  |  |  | Testing accuracy | (%) | 100 | 100 | 100 | 100 |  |  | 100 | 100 |
|  |  |  | Time (s) |  | 85.359 | 50.031 | 56.437 | 6.438 |  |  | 57.469 | 43.25 |
| Proposed | method | (systems 1 + 2) | Best features |  | F14 | F7,F10 | F18 | F5,F10 |  |  | F18 | F17,F18 |
|  |  |  | Testing accuracy | (%) | 100 | 100 | 100 | 100 |  |  | 100 | 98.36 |
|  |  |  | Time (s) |  | 38.516 | 54.861 | 34.312 | 184.921 |  |  | 46.047 | 44.578 |

Table 3

Comparison of testing accuracies and most discriminative feature subsets of each SVM and of each method.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SVM no.  S0 S1 S2 S3 S4 S5 | | | | | | | | | |
| Proposed method (systems | 1 + 2) | Best features |  | F14 | F7,F10 | F18 | F5,F10 | F18 | F17,F18 |
|  |  | Testing accuracy | (%) | 100 | 100 | 100 | 100 | 100 | 98.36 |
|  |  | Time (s) |  | 38.516 | 54.861 | 34.312 | 184.921 | 46.047 | 44.578 |
| *F*-score |  | Best features |  | F14 | F4,F10,F35,F47 | F18,F34 | F5,F9,F10 | F18 | F17,F18,F34 |
|  |  | Testing accuracy | (%) | 100 | 100 | 100 | 100 | 100 | 98.36 |
|  |  | Time (s) |  | 96.375 | 55.11 | 34.594 | 183.453 | 46.688 | 44.687 |
| *T*-test |  | Best features |  | F14 | F4,F7,F12,F10,F35,F47 | F18,F34 | F5,F10 | F18 | F17,F18,F34vsp:0.5 |
|  |  | Testing accuracy | (%) | 100 | 100 | 98.20 | 100 | 100 | 98.36 |
|  |  | Time (s) |  | 98.281 | 54.813 | 34.266 | 167.187 | 45.593 | 44.532 |
| Entropy |  | Best features |  | F14 | F6,F9,F12,F13 | F19,F38,F42, | F38,F39,F40,F42, | F42,F45,F46 | F33,F37,F38,F39,F40, |
|  |  |  |  |  |  | F43,F43,F45,F46 | F43,F44,F45,F46 |  | F41,F42,F43,F44,F45,F46 |
|  |  | Testing accuracy | (%) | 100 | 100 | 100 | 100 | 100 | 98.36 |
|  |  | Time (s) |  | 87.765 | 55.157 | 34.672 | 186.062 | 46.562 | 45.655 |
| Bhattacharyya distance |  | Best features |  | F14 | F6,F7,F9,F12,F13 | F19,F40,F42, | F38,F39,F40,F42, | F17,F18 | F33,F38,F39,F40,F41,F42, |
|  |  |  |  |  |  | F43,F44,F45,F46 | F43,F44,F45,F46 |  | F43,F44,F45,F46 |
|  |  | Testing accuracy | (%) | 100 | 100 | 100 | 100 | 100 | 98.36 |
|  |  | Time (s) |  | 84.844 | 55.187 | 34.344 | 184.515 | 46.515 | 44.547 |

Table 4

Number of features obtained by each method.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  | S0 | S1 | S2 | S3 | S4 | S5 |
| Proposed method | 1 | 2 | 1 | 2 | 1 | 2 |
| *F*-score | 1 | 4 | 2 | 3 | 1 | 3 |
| *T*-test | 1 | 6 | 2 | 2 | 1 | 3 |
| Entropy | 1 | 4 | 7 | 8 | 3 | 11 |
| Bhattacharyya distance | 1 | 5 | 7 | 8 | 2 | 10 |
|  | | | | | | |

Table 5

Comparison of the proposed method and SFFS.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SVM no.  S0 S1 S2 S3 S4 S5 | | | | | | | | | |
| Proposed | method | Best features |  | F14 | F7,F10 | F18 | F5,F10 | F18 | F17,F18 |
|  |  | Validation accuracy | (%) | 100 | 100 | 100 | 100 | 100 | 100 |
|  |  | Testing accuracy (%) |  | 100 | 100 | 100 | 100 | 100 | 98.36 |
|  |  | Time (s) |  | 277.5 | 104.2 | 67.7 | 378.9 | 79.9 | 75.6 |
| SFFS |  | Best features |  | F14 | F1,F37 | F18 | F4,F38 | F18 | F4,F17 |
|  |  | Validation accuracy | (%) | 100 | 100 | 100 | 100 | 100 | 100 |
|  |  | Testing accuracy (%) |  | 100 | 100 | 100 | 100 | 100 | 98.36 |
|  |  | Time (s) |  | 10269 | 3335.8 | 1003.3 | 1255.4 | 594 | 977.2 |

much more time than the proposed method. The speed improvement rate is calculated using (*the execution time of SFFS*)*/*(*the execution time of proposed method*) in [Table 5](#_bookmark9). The speed improvement rates of S0–S5 are 37.0, 32.0, 14.8, 3.31, 7.43, and 12.93, respectively. Accordingly, the execution time of the proposed method is about 3–37 times shorter than that for SFFS. As mentioned in Section 3, there are two adjustable weighting parameters, *wF* and *wT* , which denote the rank of the intersectional features obtained from the *F*-score and *T*-test ranking lists, respectively. To determine the appropriate values of the two adjustable weights, the range of [[1,10]](#_bookmark16) was used to select the best features of the follicle- and the ﬁbrosis-based nodules. [Fig. 6](#_bookmark10) shows the number of selected features for various values of *wF* and *wT*. From [Fig. 6](#_bookmark10), the suggested range for the two adjustable

weights is [[1,3]](#_bookmark16). This range is large enough to select the best features. The results of each branch in [Fig. 5](#_bookmark4) within this range are shown in [Fig. 7](#_bookmark11).

Using the proposed range of [[1,3]](#_bookmark16) for the adjustable weights produces 9 possible combinations for selecting the best features. From Figs. 6 and 7, many combinations of weights result in the same number of best features. To reduce the execution time, when it is known that the top *n* features between the *i*th and the (*i* + 1)th ﬁnal ranking list are the same, the (*i* + 1)th ﬁnal ranking list is not used to select features. Another consideration is the number of best features. If the number of best features equals 1, the remaining combinations are not used to select features. This explains the reduction in execution time of the proposed method compared with that of SFFS in [Table 3](#_bookmark7).

1 2 3 4 5 6 7 8 9 10

*wT*

*wF*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 4 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 5 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 6 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 7 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 8 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 9 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 |
| 10 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 |

Fig. 6. Number of best features selected from the follicle- and the ﬁbrosis-based nodules in the range of [[1,10]](#_bookmark16) for the two adjustable weights. The rows and the columns are *wT* and *wF*, respectively. The elements of this matrix are the number of the selected features.

|  |  |  |
| --- | --- | --- |
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 1 |

|  |  |  |
| --- | --- | --- |
| 2 | 2 | 2 |
| 2 | 2 | 2 |
| 3 | 2 | 2 |

|  |  |  |
| --- | --- | --- |
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 2 | 1 | 1 |

To demonstrate that the SVM can ﬁnd the optimal hyperplane to separate two classes and that the best features obtained using the proposed method have signiﬁcant discriminability, [Fig. 8](#_bookmark12) shows the distribution of each pair of classes in the feature space with the optimal hyperplane. The hyperplane segments the two classes correctly. Since S1, S3, and S5 use two features, they can be presented in two-dimensional space with a hyperplane. Only one feature is used by S0, S2, and S4, so they are presented in one- dimensional space, as shown in [Fig. 9](#_bookmark13)(a, c, and e). [Figs. 9](#_bookmark13)(b, d, and f) show the output values of the kernel function of each SVM. [Fig. 9](#_bookmark13)(a) shows the distribution of non-nodules and thyroid nodules in the F14 space, and [Fig. 9](#_bookmark13)(b) shows the output values obtained from the kernel function of S0. [Fig. 9](#_bookmark13)(c) shows the distribution of cellular and non-cellular follicles of thyroid nodules in the F18 space, and [Fig. 9](#_bookmark13)(d) shows the output values obtained from the kernel function of S2. [Fig. 9](#_bookmark13)(e) shows the distribution of two kinds of cell with follicles of thyroid nodules in the F18 space. [Fig. 9](#_bookmark13)(f) shows the output values obtained from the kernel function of S4. [Figs. 9](#_bookmark13)(b, d, and f) show that the output values of the kernel functions are separated; thus, the SVMs can correctly identify the components of a thyroid nodule using the selected features.





|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *wT*  *wF* | 1 | 2 | 3 | *wT*  *wF* | 1 | 2 | 3 | *wT*  *wF* | 1 | 2 | 3 |
| 1 |  |  |  | 1 |  |  |  | 1 |  |  |  |
| 2 |  |  |  | 2 |  |  |  | 2 |  |  |  |
| 3 |  |  |  | 3 |  |  |  | 3 |  |  |  |
|  |  | S0 |  |  |  | S1 |  |  |  | S2 |  |
| *wT wF* | 1 | 2 | 3 | *wT wF* | 1 | 2 | 3 | *wT wF* | 1 | 2 | 3 |
| 1 |  |  |  | 1 |  |  |  | 1 |  |  |  |
| 2 |  |  |  | 2 |  |  |  | 2 |  |  |  |
| 3 |  |  |  | 3 |  |  |  | 3 |  |  |  |
|  |  | S3 |  |  |  | S4 |  |  |  | S5 |  |

Fig. 7. Adjustable weights of Eq. (9) and the best choice: (a)–(f) Elements in these matrixes represent the number of best features obtained from each combination of weights. The rows and the columns are *wT* and *wF*, respectively. S0–S5 are the corresponding braches in [Fig. 6](#_bookmark10).

|  |  |  |
| --- | --- | --- |
| 3 | 3 | 3 |
| 2 | 3 | 3 |
| 2 | 2 | 3 |

|  |  |  |
| --- | --- | --- |
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 1 |

|  |  |  |
| --- | --- | --- |
| 2 | 2 | 2 |
| 2 | 2 | 2 |
| 2 | 2 | 2 |

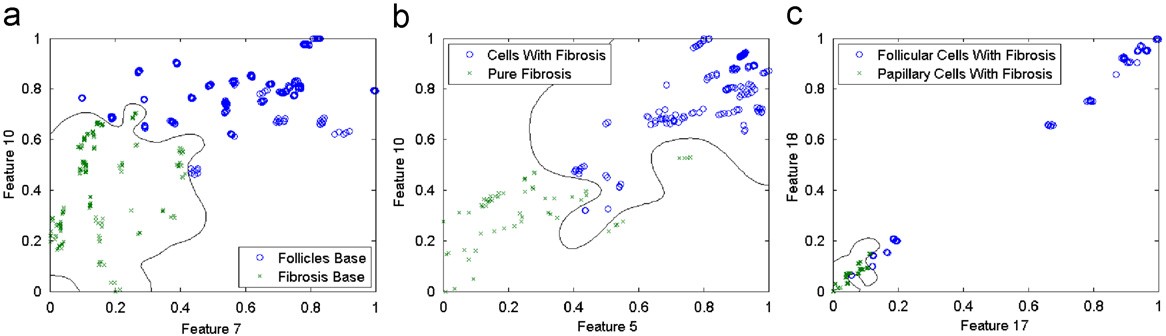


Fig. 8. Distribution of two classes of thyroid nodule with the optimal hyperplane in the feature space. The hyperplanes are shown in gray: (a) Distribution of two major bases of thyroid nodule in the F7 and F10 space. (b) Distribution of cellular and non-cellular ﬁbrosis of thyroid nodules in the F5 and F10 space. (c) Distribution of two kinds of cell with ﬁbrosis of thyroid nodules in the F17 and F18 space. The corresponding SVMs of (a)–(c) are S1, S3, and S5, respectively.

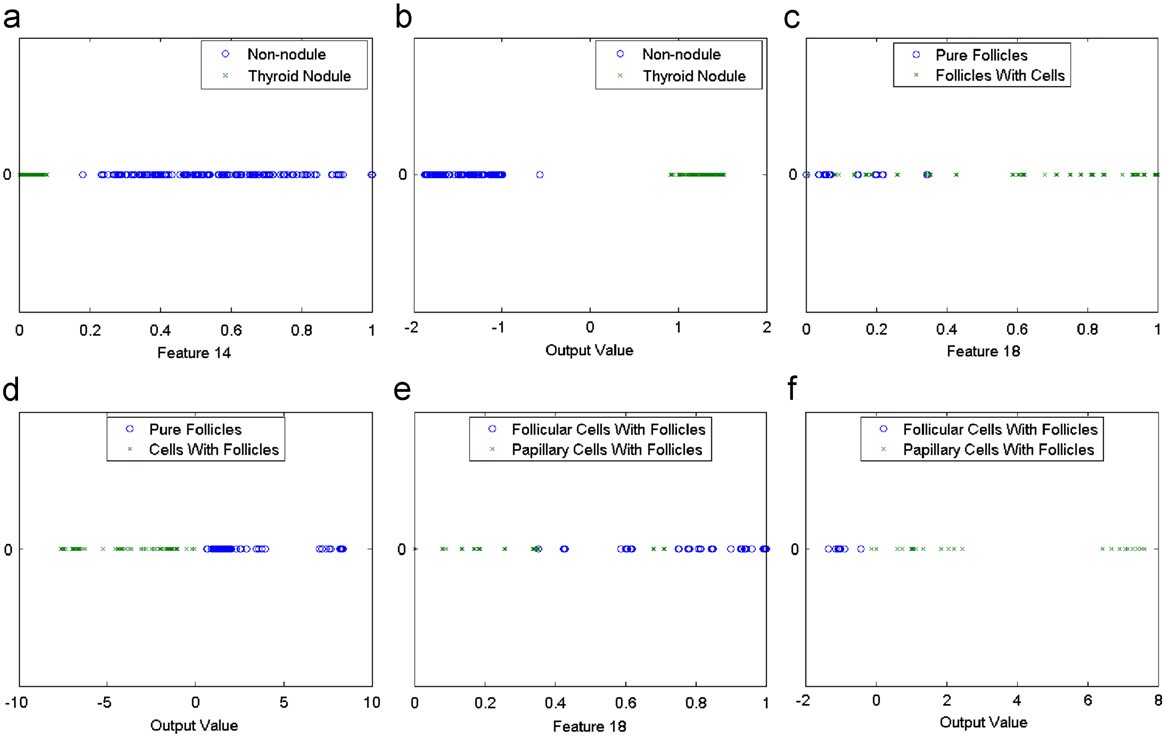


Fig. 9. Distribution of two classes of thyroid nodule in the feature space and the corresponding output values of the kernel function of SVM.

Table 6

Best settings and performance of MLP, PCA, RBF, and SOFM.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Brach  S0 | | | no. | in | Fig. | 5 | S1 | S2 | S3 | S4 | S5 |
| MLP | Processing elements | 1 | 1 | | | | | 1 | 1 | 4 | 1 |
|  | Activation function | *Tanh* | *Tanh* | | | | | *Tanh* | *Sigmoid* | *Tanh* | *Tanh* |
|  | Testing accuracy (%) | 100 | 95.90 | | | | | 74.77 | 97.56 | 94.87 | 98.36 |
| PCA | Processing elements | 1 | 2 | | | | | 1 | 1 | 1 | 1 |
|  | Activation function | *Tanh* | *Tanh* | | | | | *Tanh* | *Tanh* | *Tanh* | *Tanh* |
|  | Testing accuracy (%) | 100 | 95.90 | | | | | 70.27 | 96.34 | 87.18 | 91.80 |
| RBF | Cluster centers | 2 | 12 | | | | | 20 | 12 | 10 | 4 |
|  | Activation function | *Tanh* | *Tanh* | | | | | *Tanh* | *Sigmoid* | *Tanh* | *Tanh* |
|  | Testing accuracy (%) | 100 | 95.90 | | | | | 90.99 | 98.78 | 94.87 | 93.44 |
| SOFM | Map size (rows ~ columns) | 1\*2 | 2\*4 | | | | | 2\*3 | 1\*4 | 1\*2 | 3\*4 |
|  | Starting radius | 0 | 1 | | | | | 1 | 0 | 0 | 2 |
|  | Activation function | *Tanh* | *Tanh* | | | | | *Tanh* | *Sigmoid* | *Tanh* | *Tanh* |
|  | Testing accuracy (%) | 100 | 96.41 | | | | | 74.78 | 97.56 | 91.03 | 93.44 |
| SVM | Testing accuracy (%) | 100 | 100 | | | | | 100 | 100 | 100 | 98.36 |

* 1. *Comparison of neural networks*

To show that the SVM has the best capability of classifying thyroid nodules, the proposed SVM-based method is compared with four neural networks: the multilayer perceptron (MLP) network, the principal component analysis (PCA) network, the radial basis function (RBF) network, and the self-organizing feature map (SOFM) network [[25]](#_bookmark32). The four neural networks were simulated using NeuroSolutions. In the comparison, the features selected by the proposed method were used to train and test the neural networks.

[Table 6](#_bookmark14) shows the settings which produce the highest performance of MLP, PCA, RBF, and SOFM, respectively. The

results show that SVM, which achieved the highest accuracy of 100%, is the best choice. Although the parameters and activation functions of the four neural networks were adjusted to obtain high performance in each branch of [Fig. 5](#_bookmark4), the testing accuracies were still less than or equal to that of SVM. Among the four neural networks, only S0 and S5 in MLP have the same accuracies as those of SVM. From [Table 6](#_bookmark14), the hyperbolic activation function, *tanh*, usually results in a higher accuracy.

* 1. *Performance of thyroid nodule classiﬁcation system*

[Table 7](#_bookmark15) shows the confusion matrix of the proposed method in classifying non-thyroid nodules and the six categories of thyroid

Table 7

Confusion matrix of the proposed method in classifying non-thyroid and the six categories of thyroid nodule.

*L*

F2 : Difference entropy ¼—

X

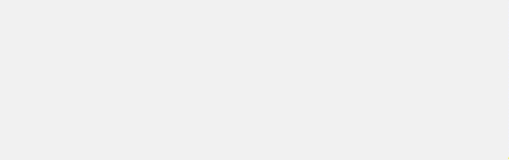
*i* ¼ 0

*px*—*y*ð*i*Þlogð*px*—*y*ð*i*ÞÞ

F3 : Difference variance ¼

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Testing accuracy | (%) | | Output  0 | class  1 | 2 | 3 | 4 | 5 | 6 |
|  | | | | | | | | | |
| Input class | | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 |
|  | | 1 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
|  | | 2 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
|  | | 3 | 0 | 0 | 0 | 100 | 0 | 0 | 0 |
|  | | 4 | 0 | 0 | 0 | 0 | 96.875 | 3.125 | 0 |
|  | | 5 | 0 | 0 | 0 | 0 | 0 | 100 | 0 |
|  | | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
|  | |  | | | | | | | |

2*L*



X

*L*

*i* ¼ 0

X

.*i*—

*L*

*j* ¼ 0

X

2

*x*—*y*ð Þ

Σ*jp j*

*px*—*y*ð*i*Þ

F4 : Sum average ¼ *ipx* þ *y*ð*i*Þ

*i* ¼ 2

2*L*

X

F5 : Sum entropy ¼— *px* þ *y*ð*i*Þlogð*px* þ *y*ð*i*ÞÞ

*i* ¼ 2

*L L*

X X

F6 : Sum of squares ¼ ð*i*—mÞ2 *p*ð*i*,*j*Þ

*i* ¼ 0 *j* ¼ 0

nodule. Non-thyroid nodules can be correctly classiﬁed by S0 with an accuracy of 100%. Only the follicular cells with ﬁbrosis cannot

be perfectly classiﬁed (accuracy of 96.875%). Other categories are

*L*

2*L*

F7 : Sum variance ¼ ð*i*—*F*5Þ2 *px* þ*y*ð*i*Þ

X

*i* ¼ 2

correctly classiﬁed with an accuracy of 100%.

*L*

X

F8 : Contrast ¼

*n* ¼ 0

2 0BX*L*

¼

X *p*ð*i*,*j*Þ1CA

1. Conclusion *L L*

*n*

@

X X

*i* 0 *j* ¼ 0

j*i*—*j*j ¼ *n*

Nodular lesions of the thyroid are very common in Taiwan.

F9 : Energy ¼ ð*p*ð*i*,*j*ÞÞ2

*i* ¼ 0 *j* ¼ 0

According to pathology, thyroid nodules are categorized as enlarged *L L*

X X

follicles, follicular cells with follicles, papillary cells with follicles,

follicular cells with ﬁbrosis, papillary cells with ﬁbrosis, and ﬁbrosis. The low resolution of sonography confuses many physicians about

F10 : Entropy ¼— *p*ð*i*,*j*Þlogð*p*ð*i*,*j*ÞÞ

*i* ¼ 0 *j* ¼ 0

the nature of various echo patterns of thyroid nodules.

To help decrease misdiagnoses, an SVM-based thyroid nodule

*L*

F11 : Local homogeneity ¼

X

*L p*ð*i*,*j*Þ

1 þð*i*—*j*Þ2

X

classiﬁcation method was proposed. 78 textural features were extracted from ROIs, which were outlined by a radiologist and

conﬁrmed by biopsy. To select the signiﬁcant features, SVMs

were used for feature selection and for obtaining the most

*i* ¼ 0 *j* ¼ 0

X*L* X*L* 3

X X

F12 : Cluster shade ¼

ð*i*—*Mx* þ*j*—*My*Þ *p*ð*i*,*j*Þ

*i* ¼ 0 *j* ¼ 0

discriminative feature set for the categories of thyroid nodule. *L L*

Each SVM was trained by the selected features of the correspond- ing category.

The experiment results show that the proposed classiﬁcation method can successfully identify six types of thyroid nodule with high accuracy. These results are very helpful for the interpretation of thyroid ultrasound and can increase diagnostic performance of ultrasound guided needle aspiration.

F13 : Cluster prominence ¼ ð*i*—*Mx* þ *j*—*My*Þ4 *p*ð*i*,*j*Þ

*i* ¼ 0 *j* ¼ 0

where *p*(*i*, *j*) is normalized GLCM, m*x* and m*y* are the mean of *px* and *py*, respectively, and s*x* and *y* are the standard deviation of *px* and *py*, respectively. `*ı* is the mean of m*x* and m*y. px*, *py*, *px+ y*, *px*—*y*, *Mx*, and *My* are deﬁned as follows:

X*L*

Acknowledgements

This work was supported by research Grants (NSC 94-2213-E- 303-002 and NSC 96-2221-E-303-001) from the National Science Council, Taiwan, and Grant (DTCRD 96(2)-13) from Buddhist Dalin

*px*ð*i*Þ¼

*py*ð*j*Þ¼

*j* ¼ 0

*L*

X

*i* ¼ 0

*p*ð*i*,*j*Þ

*p*ð*i*,*j*Þ

*L L*

X X

Tzu Chi General Hospital, Chia-Yi, Taiwan. The authors would like to thank the department of pathology, Buddhist Dalin Tzu Chi General Hospital, Chia-Yi, Taiwan, for their support and

*px* þ *y*ð*k*Þ¼ *p*ð*i*,*j*Þ, *k* ¼ 2,3, ... ,2*L*

*i* ¼ 0 *i j* ¼*j* 0 *k*

þ ¼

guidance.

*L L*

*px*—*y*ð*k*Þ¼ *p*ð*i*,*j*Þ, *k* ¼ 0,1, ... ,*L*

X

X

*i* 0 *j* ¼ 0

¼

j*i*—*j*j ¼ *k*

Appendix

Some of the features used in this paper are shown below:

*L L*

*Mx* ¼ *ip*ð*i*,*j*Þ

X

X

*i* ¼ 0 *j* ¼ 0

1. *Gray level co-occurrence matrix*:

*L*

*My* ¼

X

*L*

*jp*ð*i*,*j*Þ

X

P*L* P*L*

ð*ij*Þ*p*ð*i*,*j*Þ—m m

*i* ¼ 0 *j* ¼ 0

The normalized GLCM is calculated using

F1 : Correlation ¼

*i* ¼ 0

*j* ¼ 0 *x y*

s*x* s*y*

*p*ð*i*,*j*Þ¼ *P*ð*i*,*j*Þ=*R*

where *p*(*i*, *j*) is the original GLCM and R is the sum of total

1 X .p Σ

elements in GLCM.

1. *Statistical feature matrix*:

Imaginary ð*x*,*y*,*k*Þ¼ 8

*n* ¼ 0

7

*Ix*ð*x*,*y*,*n*Þsin

4 *kn*

To calculate the feature, the following dissimilarity matrix is required:

Dissimilarity matrixðD*x*,D*y*Þ¼ *E*fj*I*ð*x*,*y*Þ—*I*ð*x* þ D*x*,*y* þ D*y*Þjg where *I*(*x*, *y*) denotes the gray-level value of a pixel at position

(*x*, *y*) in an image *I*. D*x* and D*y* are user-speciﬁed inter-sample spacing distances, which are equal to the height and half of the width of ROI in this analysis, respectively. E{ · }denotes the

expectation operation within the texture area. When the dissimilarity matrix is obtained, the dissimilarity of ROI can be calculated using

PD*x*—1 PD*y*—1 .Dissimilarity matrixð*x*,*y*ÞΣ

Each coefﬁcient of these maps consists of two parameters:

magnitude and phase angle. They are respectively deﬁned as

*magnitude* ð*x*,*y*,*k*Þ¼ qﬃRﬃﬃeﬃﬃﬃaﬃﬃlﬃﬃﬃðﬃﬃ*x*ﬃﬃ,ﬃ*y*ﬃﬃﬃ,ﬃ*k*ﬃﬃﬃÞﬃ2ﬃﬃﬃþﬃﬃﬃﬃIﬃﬃmﬃﬃﬃﬃaﬃﬃﬃgﬃﬃiﬃﬃnﬃﬃﬃaﬃﬃrﬃﬃyﬃﬃﬃﬃðﬃﬃ*x*ﬃﬃ,ﬃ*y*ﬃﬃﬃ,ﬃ*k*ﬃﬃﬃÞﬃ2ﬃﬃ

*phase*\_*angle x*,*y*,*k* tan—1 Imaginary ð*x*,*y*,*k*Þ

Σ Σð Þ¼

Real ð*x*,*y*,*k*Þ

References

F14 : Dissimilarity ¼

*x* ¼ 0

*y* ¼ 0

D*x*D*y*

1. M.C. Frates, C.B. Benson, J.W. Charboneau, E.S. Cibas, O.H. Clark, B.G. Coleman,

J.J. Cronan, P.M. Doubilet, D.B. Evans, J.R. Goellner, I.D. Hay, B.S. Hertzberg,

C.M. Intenzo, R.B. Jeffrey, J.E. Langer, P.R. Larsen, S.J. Mandel, W.D. Middleton,

1. *Gray level run-length matrix*:

P*L*

P*C*—1 *p*ð*i*,*j*Þ=ð*j* þ1Þ2

¼

C.C. Reading, S.I. Sherman, F.N. Tessler, Management of thyroid nodules detected at US: society of radiologists in ultrasound consensus conference

F15 : Short run emphasis ¼

¼

*i* 0 *j* 0

P*L* P*C*—1 *p*ð*i*,*j*Þ

statement, Radiology 237 (2005) 794–800.

1. D.S. Cooper, G.M. Doherty, B.R. Haugen, R.T. Kloos, S.L. Lee, S.J. Mandel, E.L.

Mazzaferri, B. McIver, S.I. Sherman, R.M. Tuttle, Management guidelines for

patients with thyroid nodules and differentiated thyroid cancer: the

F16 : Long run emphasis ¼

P

*L*

*i* ¼ 0

P

*i* ¼ 0

*j* ¼ 0

P*C*—1 ð*j* þ1Þ2 *p*ð*i*,*j*Þ

American thyroid association guidelines taskforce, Thyroid 16 (2006)

109–142.

*L*

*j* ¼ 0

P ð Þ

*i* ¼ 0

P*L*

*C*—1 *p i*,*j*

*j* ¼ 0

. P*C*—1 Σ2

1. M.R. Carol, R.W. Stephanie, J.W. Charboneau, J.A. Johnson, Diagnostic Ultrasound, third ed., Elsevier Mosby, Philadelphia, PA, 2005, pp. 735–757.
2. T.C. Chang, C.T. Hong, S.L. Chang, H.C. Hsieh, K.Y. Liaw, S.W. How, Correlation

F17 : Gray level nonuniformity ¼

P

*i* ¼ 0

*L*

*i* ¼ 0

*j* ¼ 0 *p*ð*i*,*j*Þ

*C*—1 *p i*,*j*

P ð Þ

*j* ¼ 0

between sonography and pathology in thyroid diseases, Journal of the Formosan Medical Association 89 (1990) 777–783.

1. F.N. Tessler, M.E. Tublin, Thyroid sonography: current applications and future directions, American Journal of Roentgenology 173 (1999) 437–443.
2. C. Lu, T.C. Chang, Y.L. Hsiao, M.S. Kuo, Ultrasonographic ﬁndings of papillary

P*C*—1 . P*L*

P

*i* ¼ 0

ð

# 

F18 : Run length nonuniformity

*j* ¼ 0

*p i*,*j* Σ2

thyroid carcinoma and their relation to pathologic changes, Journal of the

Formosan Medical Association 93 (1994) 933–938.

F19 : Run percentage ¼

*L*

Þ

*i* ¼ 0

P

¼ *L*

*i* ¼ 0

*j* ¼ 0

*C*—1 *p i*,*j*

P ð Þ

*j* ¼ 0

*HW*

P*C*—1 *p*ð*i*,*j*Þ

1. L.S. Chow, H. Gharib, J.R. Goellner, J.A. van Heerden, Nondiagnostic thyroid

ﬁne-needle aspiration cytology: management dilemmas, Thyroid 11 (2001) 1147–1151.

1. R.M. Haralick, K. Shanugam, I. Dinstein, Textural features for image classiﬁcation, IEEE Transactions on Systems, Man and Cybernetics 3 (1973) 610–621.

where *p*(*i*, *j*) is the value located at the *i*th row and *j*th column of GLRLM.

1. *Neighboring gray level dependence matrix*:
2. L.G. Shapiro, G.C. Stockman, Computer Vision, Prentice-Hall, 2001.
3. G.V. Wouwer, P. Scheunders, D.V. Dyck, Statistical texture characterization from discrete wavelet representations, IEEE Transactions on Image Proces- sing 8 (1999) 592–598.
4. C.M. Wu, Y.C. Chen, Statistical feature matrix for texture analysis, Graphical

P*L* P*N*

.*Q k*,*s* =*s*2 Σ2

Models and Image Processing 54 (1992) 407–419.

F30 : Small number emphasis ¼

*k* ¼ 0

*s* ¼ 1 ð Þ

*R*

2

1. Y.N. Sun, M.H. Horng, X.Z. Lin, J.Y. Wang, Ultrasonic image analysis for liver diagnosis, IEEE Engineering in Medicine and Biology Magazine 15 (1996) 93–101.

F31 : Large number emphasis ¼

*L*

*k* ¼ 0

P

*N*

*s* ¼ 0

P

*R*

ð*s Q* ð*k*,*s*ÞÞ

1. M.M. Galloway, Texture analysis using gray run lengths, Computer Graphics and Image Processing 4 (1975) 172–179.
2. L.H. Siew, R.M. Hodgson, E.J. Wood, Texture measures for carpet wear

P*L* . P*N* Σ2

assessment, IEEE Transactions on Pattern Analysis and Machine Intelligence

F32 : Number nonuniformity ¼

P*L*

*k* ¼ 0

P*N*

*s* ¼ 0 *Q* ð*k*,*s*Þ

*R*

ð*Q* ð*k*,*s*ÞÞ

2

10 (1988) 92–105.

1. C.J. Sun, G.W. Wee, Neighboring gray level dependence matrix for texture

classiﬁcation, Computer Vision, Graphics, and Image Processing 23 (1983) 341–352.

F33 : Sencond moment ¼

P—

P

*k* ¼ 0

*s* ¼ 0

*R*

1. M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, Image coding using

wavelet transform, IEEE Transactions on Image Processing 1 (1992) 205–220.

F34 : Entropy ¼

*L*

*k* ¼ 0

*N*

*s* ¼ 0

*Q* ð*k*,*s*Þlogð*Q* ð*k*,*s*ÞÞ *R*

1. F. Zhou, J.F. Feng, Q.Y. Shi, Texture feature based on local Fourier transform, Proceeding of the International Conference on Image Processing 2 (2001) 610–613.

where *Q*(*k*, *s*) is NGLDM, and *R* is equal to (*H*— 2*d*) ~ (*W* — 2*d*).

1. *Fourier feature based on local Fourier coefﬁcients*:

The following equations can be used to calculate the real and imaginary parts of *F*(*x*, *y*, *k*).

*F*ð*x*,*y*,*k*Þ¼ 1 X *I*ð*x*,*y*,*n*Þ*ej*p*kn*

7

1. C. Cortes, V. Vapnik, Support-vector network, Machine Learning 20 (1995) 273–297.
2. Y.W. Chen, C.J. Lin, Combining SVMs with Various Feature Selection Strategies, Feature Extraction, Foundations and Applications, Springer, 2006.
3. C.L. Huang, H.C. Liao, M.C. Chen, Prediction model building and feature selection with support vector machines in breast cancer diagnosis, Expert Systems with Applications 34 (2008) 578–587.

8

*n* ¼ 0

1 X .

7

4

.p Σ

.p ΣΣ

1. Sergios Theodoridis, Konstantinos Koutroumbas, Pattern Recognition, second ed., Elsevier, 2003.
2. P. Pudil, F.J. Ferri, J. Novovicova, J. Kittler, Floating search methods for feature

¼ 8

*n* ¼ 0

*I*ð*x*,*y*,*n*Þ cos

1 X

7

4 *kn*

þ *j* sin

.p Σ

4 *kn*

, 0 r *k*r 7

selection with nonmonotonic criterion functions, in: Proceedings of the 12th IAPR International Conference on Pattern Recognition, vol. 2, October 1994.

1. LOGIQ 700 Expert Series: Conformance Statement for DICOM V3.0. Direction

Real ð*x*,*y*,*n*Þ¼ 8

*n* ¼ 0

*I* ð*x*,*y*,*n*Þcos

4 *kn*

2243479-100, Revision 1. Technical Publications. GE Healthcare, Chalfant St.

Giles, UK. Available at: /[http://www.gehealthcare.com/usen/interoperabil](http://www.gehealthcare.com/usen/interoperability/dicom/docs/logiq700_rev2.pdf) [ity/dicom/docs/logiq700\_rev2.pdf](http://www.gehealthcare.com/usen/interoperability/dicom/docs/logiq700_rev2.pdf)S.

1. C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, 2001. Software available at: /[http://www.csie.ntu.edu.tw/](http://www.csie.ntu.edu.tw/%7Ecjlin/libsvm) ~ [cjlin/libsvm](http://www.csie.ntu.edu.tw/%7Ecjlin/libsvm)S.
2. S. Haykin, Neural Networks: A Comprehensive Foundation, second ed., Prentice-Hall, 1999.
3. C.J. Lin, A formal analysis of stopping criteria of decomposition methods for support vector machines, IEEE Transactions on Neural Networks 13 (2002) 1045–1052.
4. S. Tsantis, D. Cavouras, I. Kalatzis, N. Piliouras, N. Dimitropoulos, G. Nikiforidis, Development of a support vector machine-based image analysis system for assessing the thyroid nodule malignancy risk on ultrasound, Ultrasound in Medicine and Biology 31 (2005) 1451–1459.
5. J.X. Wang, J.K. Yu, L. Wang, Q. Liu, J. Zhang, S. Zheng, Application of serum protein ﬁngerprint in diagnosis of papillary thyroid carcinoma, Proteomics—Clinical Applications 6 (2006) 5344–5349.

Chuan-Yu Chang received the BS and MS degrees in nautical technology and electrical engineering from National Taiwan Ocean University, Keelung, Taiwan, in 1993 and 1995, respectively, and the Ph.D. degree in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan, in 2000. From 2001 to 2002, he was with the Department of Computer Science and Information Engineering, Shu-Te University, Kaohsiung, Taiwan, ROC. Since 2003 he has been with the Institute of Computer Science and Information Engineering, National Yunlin University of Science and Technology, where he is currently an associate professor. His current research interests include neural networks, and their applications to medical image processing, wafer defect inspection, digital watermarking, face detection, and face recognition. In the above areas, he has over 100 publications in journals and conference proceedings.

Dr. Chang received the excellent paper award of the Image Processing and Pattern Recognition society of Taiwan in 1999, 2001, and 2009. He also was the recipient of the best paper award in International Computer Symposium in 1998 and the best paper award in the Conference on Artiﬁcial Intelligence and Applications in 2001, 2006, 2007, and 2008, and the best paper award in National Computer Symposium in 2005, respectively. Dr. Chang is a senior member of IEEE, the IPPR (Chinese Image Processing and Pattern Recognition Society), and the TAAI (Taiwanese Association for Artiﬁcial Intelligence), and is listed in *Who’s Who in Science and Engineering*, *Who’s Who in the World, Who’s Who in Asia, and Who’s Who of Emerging Leaders*.

Shao-Jer Chen received his M.D. degree from Kaohsiung Medical University, Kaohsiung, Taiwan, in 1991. He completed a radiology residency at National Cheng Kung University Hospital (1996). He also received his Ph.D. degree in Biomedical Engineering from National Cheng Kung University, Tainan, Taiwan, in 2008. Now he is the visiting staff of the Department of Medical Imaging at the Buddhist Dalin Tzu Chi General Hospital, Chia-Yi, Taiwan and the assistant professor at the School of Medicine, Buddhist Tzu Chi University, Hualien, Taiwan. His main research interests include medical imaging analysis, ultrasound, computer tomography and magnetic resonance imaging.

Ming-Fong Tsai received his M.S. degrees in Computer Science and Information Engineering from National Yunlin University of Science and Technology, Yunlin, Taiwan, in 2008. His interest is in medical image processing. He is currently an R.D. Engineer in GIPS Technology Co., Ltd, Taiwan.