Classification Algorithms for Quantitative Tissue Characterization of Diffuse Liver Disease from Ultrasound Images

Yasser M. Kadah, Aly A. Farag,* *Member, IEEE,* Jacek M. Zurada, *Fellow, IEEE,*
Ahmed M. Badawi, and Abou-Bakr M. Youssef

Abstract—Visual criteria for diagnosing diffused liver diseases from ultrasound images can be assisted by computerized tissue classification. Feature extraction algorithms are proposed in this paper to extract the tissue characterization parameters from liver images. The resulting parameter set is further processed to obtain the minimum number of parameters which represent the most discriminating pattern space for classification. This preprocessing step has been applied to over 120 distinct pathology-investigated cases to obtain the learning data for classification. The extracted features are divided into independent training and test sets, and are used to develop and compare both statistical and neural classifiers. The optimal criteria for these classifiers are set to have minimum classification error, ease of implementation and learning, and the flexibility for future modifications. Various algorithms of classification based on statistical and neural network methods are presented and tested. We show that very good diagnostic rates can be obtained using unconventional classifiers trained on actual patient data.

I. INTRODUCTION

THE use of ultrasonography as an imaging modality has become widely spread because of its ability to visualize main organs with no deleterious effects. The basic idea of ultrasonic imaging is to send a fine beam of ultrasonic waves through the human tissues and then receive the characteristic echo reflections from the internal body structures to form the ultrasound image. The different gray levels of this image represent the acoustic properties of the human tissues such as attenuation of acoustic waves, speed of sound, and acoustic impedance of the different body structures. All these factors contribute to the shape and intensity of the returned waves according to the underlying tissue properties and hence, this fact is the basis for the use of ultrasonography as an imaging technique. The main limitation for ultrasound is its inherent inability to visualize air-containing or bony structures. This

limitation does not apply for most of the abdominal body structures as they are composed of soft tissues and blood vessels [1].

Liver diseases are taken seriously because of the liver's vital importance to the life of the patient. Liver pathologies can be classified into two main categories according to the degree of dispersion of the disease. The first category is the *localized liver diseases* in which the pathology is concentrated in small spot(s) in one or both of the liver lobes while the rest of the liver tissue remains normal. The second category is the *diffused liver diseases* in which at least one complete lobe of the liver is affected by the disease or, in other words, the disease is distributed over the whole liver volume. This classification does not imply that the second category is a later stage of the first or that it is more serious. Both categories encounter benign and malignant types of diseases and should be treated in totally different ways from each other [1], [2], [5].

Visual criteria for diagnosing diffused liver diseases are in general confusing and highly subjective. They depend on the ability of the sonographer to observe certain textural characteristics from the image and to compare them with those developed for different pathologies in order to determine the type of the disease. Examples for these features are *texture homogeneity* and *texture echogenicity* where their description can be widely debated among experienced sonographers especially in marginal cases. Moreover, some diseases are highly similar in their diagnostic criteria, which tend to confuse the sonographers even more. Diagnostic accuracy using only visual interpretation is currently estimated to be around 72% [15]–[17]. Examples of these cases are illustrated in Figs. 1–3. Visual examination of these images does not produce conclusive diagnosis. Therefore, physicians may resort to invasive methods such as the pathology investigation of ultrasonically guided needle biopsies. Although this technique is considered to be the golden test for diagnosis in terms of accuracy, it has the disadvantage of being invasive and more importantly, it poses a risk of cancer spreading if it cuts through a localized cancer area [2], [8].

To solve some of these problems in diagnosis, researchers have developed more quantitative criteria using computer-based systems (e.g., [2], [3], [13], [14], [28]–[35]). Good results have been reported for the thyroid [31], breast [13], and liver [2], [32], [35]. Garra *et al.* [34] compared the performance of computer-based tissue characterization systems to that of
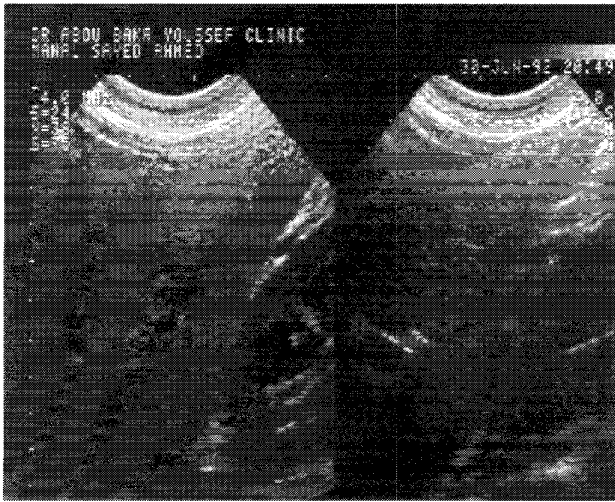
Fig. 1. Liver B-Scan ultrasound images for a normal case showing a homogeneous texture of a liver lobe and a part of the right kidney. The high echoes in the first part of the image are due to reverberation.
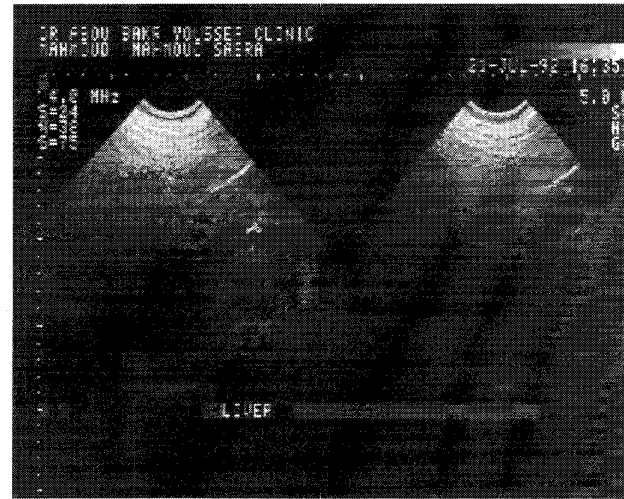

Fig. 3. Ultrasound scans for a fatty case with different machine settings. Different settings produce different image statistics, therefore this image can not be compared to the images in Figs. 1 and 2 using the developed technique.
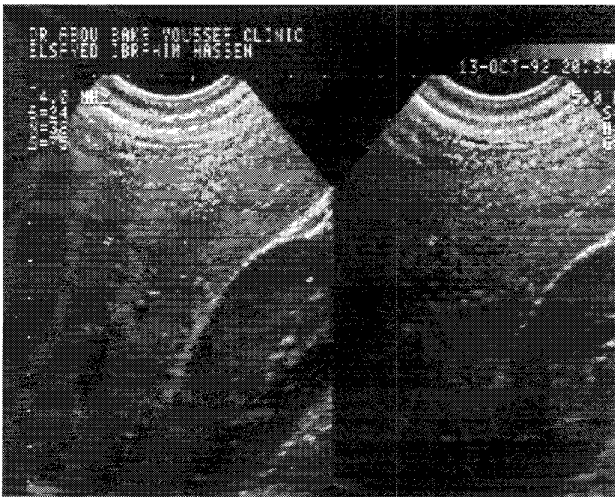

Fig. 2. Ultrasound scans for a cirrhotic case. It is difficult to distinguish this case from normal.

human observer on several types of diffuse liver disease. The results suggested that the use of quantitative tissue characterization could significantly increase the usefulness of ultrasound for the evaluation of diffuse liver disease. These results clearly show that the new approach, which can be denoted as quantitative tissue characterization technique (QTCT), is gaining more acceptance and appreciation from the ultrasound diagnosis community. It has the potential to significantly assist sonographers to achieve better diagnostic rates.

Generally speaking, QTCT is based on extracting parameters from the returned ultrasound echoes for the purpose of identifying the type of tissue present in the ultrasound scan plane. These parameters can be divided into two main categories according to their origin.

RF Signal   Extracted from the returned RF echoes prior to machine processing (e.g., attenu-

ation [4], [36] and backscattering parameters [37]).

Image Texture   Extracted from the video image after the echo processing is performed in the machine. Such parameters include the statistical characteristics of the gray level distribution in a certain region of interest (ROI) in the image. For example, parameters obtained from the image histogram, image gradient, co-occurrence matrices, and run-length matrix [3].

The first type has the advantage of being free from any machine processing distortions, while the second has the advantage of being easier to implement. Using the above parameters, QTCT attempts to simulate the process of diagnosis inside the human brain by using a numerical classifier. For the case of textural parameters, there is an explicit correspondence between the values of the parameters and the visual appearance of the ultrasound image. For example, the values of the gray level histogram parameters depend mainly on the intensity of the ROI, while the values of the gray level co-occurrence matrix parameters depend on the contrast/homogeneity of the image. These parameters will be formally defined in Section III-A.

In a classical statistical classification problem, a measurement vector is assigned to the most probable class among those used in the classifier design. A class is a collection of elements having similar properties. For example, in diffuse liver disease the possible classes can be normal, fatty, and cirrhotic. However, subclasses inside each of these classes may also be defined to correspond to different stages of the disease. A symbolic representation of this problem is shown in Fig. 4, where each cluster represents the volume in the decision space assigned to a particular class. In general, these clusters may overlap and create regions of classification uncertainty. As a result, the efficiency of the process of classification, and hence
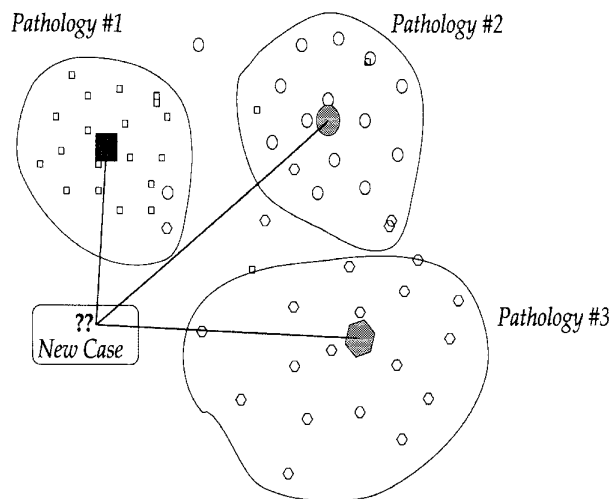
Fig. 4.   Illustration of the idea of diagnosis in the parameter space. Each cluster corresponds to a particular pathological condition and the new case is required to be optimally classified as a member of one of these clusters.

that of QTCT, depends mainly on minimizing errors due to these overlaps. This can be achieved by optimizing the process of parameter selection and the choice of the classifier.

Several classification methods have been used for QTCT (e.g., linear discriminant and multivariate analysis [37] and clustering analysis [2]). Nevertheless, attempts to improve QTCT have been focused mainly on selection of the parameters for the classifiers, while the issue of optimal classifier selection has been largely ignored in the ultrasound literature. In [20] and [21], we investigated several efficient statistical- and neural-based techniques for noninvasive diffuse liver disease tissue classification. In this paper, we present a detailed description of the algorithms and the results of applying these techniques on actual patient data. Our main goal is to investigate the use of well-known quantitative classification techniques in QTCT, and evaluate their relative performance in liver tissue characterization. The paper also introduces a novel classification algorithm based on the functional link neural networks concept.

II. DATA ACQUISITION

The ultrasound images used in this research were obtained by a Kretz-320 mechanical sector scanner ultrasound machine. A Matrox PIP-512 frame grabber card on an IBM-PC was used to capture 512×512 image frames with gray level resolution of 8 b/pixel. Images were stored in a special frame buffer that can be manipulated by the programmer through a special interface library. A number of software tools were developed on this system to allow the sonographer to define the ROI in the image for further analysis. A block diagram of the system with the interconnections between its different parts is provided in Fig. 5.

To obtain reproducible results, a preparatory calibration procedure was conducted before the data collection phase. During this procedure, images of excised fresh animal liver were

scanned. The procedure of choosing the ROI and parameter calculation was performed under different imaging conditions such as scanning frequency, machine settings (TGC, gain, contrast, and zooming), in addition to varying the size and location of the ROI. The results of this study, in addition to research works in [35], [38], showed that the following scanning conditions must be standardized for all scans to ensure the fidelity of the tissue characterization procedures:

1) Ultrasound machine settings (e.g., TGC, FOCUS, and ZOOM controls) which change the overall image gain and produce uncontrollable effects. It is important that the same settings are used for all tissues used in an experiment, to avoid deviations in the image statistics. Also, the transducer type should be the same for all cases to avoid any bias between different interpolation schemes used for different scan protocols. Moreover, the frequency of ultrasound waves used must be the same since the attenuation of ultrasonic waves depends mainly on this frequency (see Fig. 3). In our case, the transducer used for all cases was a 4-MHz mechanical sector scanner.

2) ROI size and shape. To obtain reliable statistics, the number of pixels in the ROI must be at least 800 pixels to provide the suitable sample size condition for reliable statistics. Also, a square shape of the region should be maintained during all procedures. This corresponds to a region that is approximately 1 cm × 1 cm with a scan resolution of 30 pixels/cm leading to the given number. The actual regions taken were 2 cm × 2 cm, which lead to a region of interest of 3600 pixels that is far beyond this minimum. These results were obtained from a calibration procedure discussed in [15], [18], and [19], and also agree with the known results in the statistical pattern recognition literature (e.g., [24]).

3) ROI location. To avoid the distorting effects in ultrasonic wave patterns, such as side lobes and grating lobes, the selected region should be selected each time along the center line of the image. Also, the depth of the ROI should be chosen such that the distorting effects of the reverberations in the shallow parts and the attenuation in the deep parts are avoided.

4) Fasting condition of the patient. It has been suggested that patients should be fasting for eight hours before any scan to avoid the effects of changing the liver glycogen and water storage on ultrasound attenuation [36]. This particular issue may not be too critical due to the sound differences between patients having the same pathological condition.

The effect of the abdominal wall thickness and composition was not considered here due to its extreme difficulty to assess and control (e.g., [39]).

Data acquisition for this research was obtained from ultrasound images in a transverse subcostal section taken for patients just preceding a needle biopsy procedure on their livers. The data from the pathology laboratory were used to identify the exact condition of all obtained images. This procedure was performed for over 120 distinct cases. The
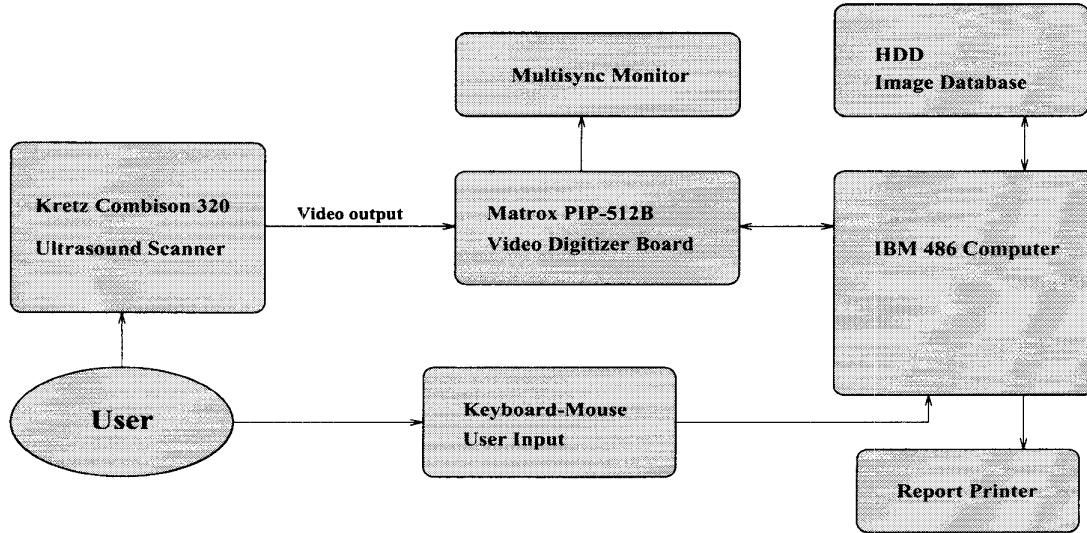
Fig. 5. Block diagram of the developed QTCT system.

disease prevalence of the three liver conditions (normal, fatty, and cirrhotic) were found to be practically similar.

During the course of data collection, a large set of statistical parameters were calculated from selected regions in each image having the standardized properties mentioned above. This parameter set was further processed using correlation measurements [19] to discard the dependent parameters to identify those which most strongly correlate to different pathologies. In our study, the final parameter set includes eight parameters: two from the image histogram (mean gray level and first-percentile), four from the gray level co-occurrence matrix (contrast, entropy, correlation, and angular second moment), in addition to attenuation estimate [4], [5] and scatterer separation distance [7]. Specific definitions of these parameters are given in Section III-A. The data set of the above parameters, collected from the 120 images corresponding to 120 patients of known pathology, was divided into two sets of equal size; training set and test set. The training set is used to derive the classifier parameters, while the test set is used to obtain the success rate of the classifier. This procedure is very important to provide bias-free statistical results. Finally, all data samples were labeled by one of three pathologies: normal, cirrhotic, or fatty liver diseases; thus, three clusters in a quantitative classification space were formed.

III. DATA ANALYSIS METHODS

A. Parameter Definitions

1) First-Order Gray Level Parameters: In this category, the parameters are derived from the gray level histogram. They describe the first-order gray level distribution without considering spatial interdependence. As a result, they can only describe the echogenicity of texture and the diffuse variation characteristics within the ROI. The two selected parameters from this category are the following.

a) The mean gray level (g_{ave})

$$g_{ave} = \frac{1}{N} \sum_{(i,j) \in \text{ROI}} g(i, j) \tag{1}$$

where $g(i, j)$ is the gray level at pixel (i, j), and N is the total number of pixels inside the ROI (i.e., N is equal to the cardinality of the ROI).

b) The first percentile of the gray level distribution (P_1)

$$\sum_{j=0}^{P_1 - 1} h_j < \frac{1}{10}$$

$$\leq \sum_{i=0}^{P_1} h_i \tag{2}$$

where h_j is the value of the gray level histogram at gray level j. In other words, the first percentile represents the gray level at or below which lies 10% of the total number of pixels inside the ROI.

2) Second-Order Gray Level Parameters: As one might guess from the definition of first-order parameters, this category of parameters describes the gray level spatial inter-relationships and hence, represent efficient measures of the gray level texture homogeneity. These parameters can be derived using several approaches such as first-order gradient distribution, gray level co-occurrence matrix, edge co-occurrence matrix, or run-length matrix. The parameters used in this work from this category are the gray level co-occurrence matrix parameters. The formal definition of this matrix is as follows:

$$Co(i, j) = \frac{1}{N} \text{cardinality} \{ [(k, l), (m, n)] \in \text{ROI}:$$
$$|k - m| = dx, \ |l - n| = dy,$$
$$\text{sign} [(k - l) \cdot (l - n)] = \text{sign} (dx \cdot dy),$$
$$g(k, l) = i, \ g(m, n) = j \} \tag{3}$$

where $Co(i, j)$ is the gray level co-occurrence matrix entry at gray levels i and j, $g(i, j)$ is the gray level of the pixel (i, j) in the ROI, N is the total number of pixels inside the ROI, and (dx, dy) is a prescribed neighborhood definition, taken in our case to be $(0, 4)$, representing an axial neighborhood definition. In other words, the entry (i, j) of this matrix describes how often the two gray levels i and j are neighbors under the given neighborhood definition. Note that this definition does not discriminate between negative and positive shifts and hence, the co-occurrence matrix is expected to be symmetric using this definition. Four parameters were derived from this matrix and are defined as follows:

a) Contrast (CON)

$$CON = \sum_{i,j \in G} (i - j)^2 \cdot Co(i, j). \qquad (4)$$

b) Angular second moment (ASM)

$$ASM = \sum_{i,j \in G} [Co(i, j)]^2. \qquad (5)$$

c) Entropy (ENT)

$$ENT = -\sum_{i,j \in G} Co(i, j) \cdot \log[Co(i, j)]. \qquad (6)$$

d) Correlation (COR)

$$COR = \frac{\sum_{i,j \in G} ijCo(i, j) - m_x \cdot m_y}{S_x \cdot S_y} \qquad (7)$$

where

$$m_x = \sum_i i \sum_j Co(i, j) \qquad (8a)$$

$$m_y = \sum_j j \sum_i Co(i, j) \qquad (8b)$$

$$S_x^2 = \sum_i i^2 \sum_j Co(i, j) - m_x^2 \qquad (8c)$$

$$S_y^2 = \sum_j j^2 \sum_i Co(i, j) - m_y^2 \qquad (8d)$$

and G is the set of available gray level values from the video digitizer.

3) Attenuation and Backscattering Parameters: We assume a simple exponential model for the backscattering echo amplitude $E(x)$

$$E(x) = \frac{E_o}{x} \sigma(x) \cdot \exp\left[-2 \int_0^x \alpha(r)\, dr\right] \qquad (9)$$

where $\sigma(x)$ represents the scattering coefficient, x is the distance to the transducer, α is the attenuation coefficient, and E_o is a constant proportionality factor. The factor of two in the exponent arises from the round trip to the reflector and back to the transducer. By sampling both sides of this formula

$$E_{ij} = \frac{E_o}{x_{ij}} \sigma_{ij} \cdot \exp\left[-2 \sum_{k=1}^{i} \sum_{l=1}^{j} \alpha_{kl}\right] \qquad (10)$$
$$(k,l) \in R_i$$

where E_{ij} is the amplitude of the echo received from pixel (i, j), σ_{ij} is the backscattering coefficient, α_{ij} is the attenuation coefficient, and R_i is the set of image coordinates along the ultrasound ray of interest. It is obvious that the summation in the exponent is performed along an ultrasound ray. From B-scan images, we can obtain values for the left side of the above formula for any given value of x_{ij} and the goal is to find α_{ij}. Several estimation procedures can be used to obtain the attenuation coefficients from the B-mode data. Examples of such techniques are as follows.

a) Maximum likelihood (ML) approach where the distribution of the echo amplitudes is assumed to follow a Rayleigh distribution [7]. Recent reports have shown that in order to develop a description of the speckle pattern, a constant level of spatially distributed specular scattering should be considered; the latter obeying a Rician probability density function [8].

b) Least squares approach after taking the log of both sides of the sampled formula.

c) Extended Prony method which consists of fitting exponentials corresponding to the original data [9], [10]. This method provided the smallest error which was estimated to be around 5% for a region of $2 \times 2 cm^2$. We used this method to derive the attenuation coefficient in this study.

4) Backscattering Coefficient: The elastic and collagen fibers exhibit an elasticity constant that is greater than that of other tissues [11]. Because of the differences in elasticity at collagenous interfaces, this will lead to increasing the scattered intensity at these sites. Hence, measuring the backscattering coefficient can provide valuable information about the structure of soft tissues since the increase of stromal density (collagen) within a tissue will increase the backscattering coefficient accordingly. The basic expression for backscattering is given by [12]

$$I(180°) = \frac{I_o V K_o^3 \gamma_k^2}{\pi R_d^2} \int_0^d N(r) R \sin(2K_o r)\, dr \qquad (11)$$

where I_o is the incident intensity, K_o is the incident wave vector, γ_k is the rms deviation of the compressibility of inhomogeneities from the mean value expressed as a fraction of that mean value, $N(r)$ is the correlation function of the scatterer (characteristic separation in scattering structures), V is the scattering volume, and R_d is the distance from the scattering volume. For a Gaussian characteristic separation

$$N(r) = \exp\left(-\frac{r^2}{\overline{a}^2}\right) \qquad (12)$$

where $\overline{a}$ is the average scatterer separation. Assuming a periodicity about the dimension d or zero scattering outside the volume of dimension d, the limits of integration can be changed to infinity instead of d, we obtain

$$\mu_d(180°) = I(180°) \frac{R_d^2}{I_o V}$$
$$= \frac{\gamma_k^2 K_o^4 \overline{a}^3}{4\sqrt{\pi}} \exp(-K_o^3 \overline{a}^3). \qquad (13)$$

For an exponential characteristic separation, we have

$$N(r) = \exp \left(-\frac{R}{a} \right). \tag{14}$$

Using the same periodicity assumption as before, we get $\mu_d(180°)$

$$\mu_d(180°) = \frac{\gamma_k^2 K_o^4 \bar{a}^3}{\pi(1 + 4K_o^2\bar{a}^2)^2}. \tag{15}$$

We used the above formula to obtain the backscattering coefficient.

B. Statistical Classifiers

Statistical classification can be divided into parametric and nonparametric techniques. Parametric statistical pattern recognition uses given or assumed information about the prior probabilities to obtain the classification. For example, in the Bayes minimum-error classifier, one can assume a Gaussian *a priori* conditional probability densities for the classification process, and estimate the parameters of the Gaussian probability density function (i.e., the mean, covariance, and correlation coefficient) from the given samples. On the other hand, the nonparametric approach does not require any *a priori* information about the probability distributions of the data in each class; classification is performed based on the provided data samples with known class memberships. For example, in the voting k-nearest neighbor technique, classification is made using the labeled design set only [24]. Hence, this approach is less subjected to the bias effects due to incorrect assumptions or parameter changes than the former [23].

An important initial step in both techniques is to divide the data set into two independent subsets, design and test sets. This preliminary step effectively avoids introducing false-negative bias effects to the error estimates. It can be shown that the error obtained by classifying an independent set of known classes is unbiased and consistent estimate for the actual error value. Also, the effect of normalization of all parameter values within a fixed range around the zero (e.g., between ±1) is studied as a possible convenient preprocessing step for proper weighing of parameters in distance-dependent measurements [23].

Before we describe the classifiers that were implemented in this study, we will define a set of parameters that will be referred to frequently with reference to statistical and neural classifiers.

False-negative rate	The probability that the classification result indicates a normal liver while the true diagnosis is indeed a liver disease (i.e., positive). This case should be completely avoided since it represents a danger to the patient.
False-positive rate	The probability that the classification result indicates a liver disease while the true diagnosis is indeed a normal liver (i.e., negative). This case can be tolerated, but should be as infrequent as possible.
Sensitivity	The conditional probability of detecting a disease while there is in fact a liver disease. By definition, Sensitivity = 1 − false-negative rate.
Specificity	The conditional probability of detecting a normal liver while the liver is indeed normal. By definition, Specificity = 1 − false-positive rate.

1) Minimum Distance Classifier: This method assumes that the classes are similar in distribution and are linearly separable. Hence, the decision lines are allocated half-way between the centers of clusters of different classes.

Algorithm 1:

Step 1) Group the design set into three supervised clusters according to their labels, representing the three pathologies of interest.

Step 2) Estimate the sample mean for each class by averaging the eight-dimensional (8-D) parameter set of the class.

Step 3) A test sample is classified by assigning it to the class which has the nearest mean vector.

Step 4) Error rate is estimated by the percentage of misclassified samples.

2) Bayes Minimum-Error Classifier: The Bayes decision rule classifies an observation (test sample) to the class that has the highest *a posteriori* probability among the three classes [23]–[25]. In this study, the data set is assumed to have a Gaussian conditional density function and the *a priori* probabilities are assumed to be equal for the three pathologies. That is

$$f_{\mathbf{X}}(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{N/2}|\mathbf{\Sigma}_i|^{1/2}}$$
$$\cdot \exp \left[-\frac{1}{2}(\mathbf{x} - \mathbf{M}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{x} - \mathbf{M}_i) \right] \tag{16}$$

and

$$P(\omega_i) = \tfrac{1}{3}, \quad i \in \{1, 2, 3\} \tag{17}$$

where

$\mathbf{x}$ 8×1 data sample from the random vector $\mathbf{X}$;
$\mathbf{M}_i$ 8×1 vector representing the sample mean of class i;
$\mathbf{\Sigma}_i$ 8×8 matrix representing the covariance matrix of class i;
ω_i class i.

The Bayes decision rule is: Choose class $j \in \{1, 2, 3\}$ if

$$f_{\mathbf{X}}(\mathbf{x}|\omega_j)P(\omega_j) = \max\{f_X(\mathbf{x}|\omega_i)P(\omega_i)|i = 1, 2, 3\}. \tag{18}$$

Since the covariance matrices are expected to be different for each class, (18) is, in general, quadratic. This equation can be evaluated numerically. However, a simpler approach is obtained through the approximation of the covariance matrices by their corresponding Toeplitz forms (e.g., [24]). Assume that each covariance matrix can be put in the form

$$\mathbf{\Sigma}_i \approx \mathbf{\Gamma}_i \mathbf{R}_i \mathbf{\Gamma}_i \tag{19}$$

$$\mathbf{\Gamma}_i = \begin{bmatrix} \sigma_{i1} & 0 & \cdots & 0 \\ 0 & \cdot & & \vdots \\ \vdots & & \cdot & 0 \\ 0 & \cdots & 0 & \sigma_{iN} \end{bmatrix} \quad (20)$$

$$\mathbf{R}_i = \begin{bmatrix} 1 & \rho_i & \cdots & \rho_i^{N-1} \\ \rho_i & 1 & & \vdots \\ \vdots & & \cdot & \rho_i \\ \rho_i^{N-1} & \cdots & \rho_i & 1 \end{bmatrix} \quad (21)$$

where σ_{ik} is the standard deviation of the kth parameter in class i and ρ_i is a fixed parameter for each class.

Notice that $\mathbf{\Gamma}_i$ is a diagonal matrix and $\mathbf{R}_i$ is Toeplitz. The parameter σ_{ik} is the standard deviation of the random variable $X_k \in \mathbf{X}$, $k \in [1, N]$; it is equal to the square root of the diagonal elements of the matrix $\mathbf{\Sigma}_i$. The parameter ρ_i is estimated by averaging those elements in the autocorrelation matrix $\mathbf{R}_i = \{R_{m,n}, \quad 1 \le m, n \le N\}$ for which $|m - n| = 1$. It is straight forward to show that

$$|\mathbf{R}_i| = (1 - \rho_i^2)^{N-1} \rightarrow |\mathbf{\Sigma}_i| \approx (1 - \rho_i^2)^{N-1} \prod_{k=1}^{N} \sigma_{ik}^2 \quad (22)$$

$$\mathbf{R}_i^{-1} = \frac{1}{1 - \rho_i^2} \begin{bmatrix} 1 & -\rho_i & 0 & \cdots & 0 \\ -\rho_i & 1 + \rho_i^2 & \cdot & \cdot & \vdots \\ 0 & \cdot & \cdot & \cdot & 0 \\ \vdots & \cdot & \cdot & 1 + \rho_i^2 & -\rho_i \\ 0 & \cdots & 0 & -\rho_i & 1 \end{bmatrix} \quad (23)$$

$$\mathbf{\Gamma}_i^{-1} = \begin{bmatrix} \frac{1}{\sigma_{i1}} & 0 & \cdots & 0 \\ 0 & \cdot & & \vdots \\ \vdots & & \cdot & 0 \\ 0 & \cdots & 0 & \frac{1}{\sigma_{iN}} \end{bmatrix} \quad (24)$$

and

$$\mathbf{\Sigma}_i^{-1} = \mathbf{\Gamma}_i^{-1} \mathbf{R}_i^{-1} \mathbf{\Gamma}_i^{-1}. \quad (25)$$

The above approximation makes it simpler to evaluate $\mathbf{\Sigma}_i^{-1}$. The Bayes minimum error classifier in (18) is implemented as follows.

Algorithm 2:

Step 1) Estimate the Gaussian distribution parameters (i.e., 8-D mean vectors and covariance matrices, $N = 8$ for our data) from the samples of each class.

Step 2) Use the Toeplitz approximation in (19)–(25).

Step 3) Substitute into the multidimensional Gaussian formulas for the three classes and multiply each of the obtained values by the *a priori* probability of its class.

Step 4) Classify a test sample by assigning it to the class having the highest *a posteriori* probability among the three classes.

Step 5) Estimate the error rate by comparing the classification results to the actual class membership.

3) Voting k-Nearest Neighbor (k-NN) Classifier: This technique is nonparametric, it assigns a test sample to the class of the majority of its k-neighbors [24]; that is, assuming that the number of voting neighbors is $k = k_1 + k_2 + k_3$ (where k_i is the number of samples from class i in the k-sample neighborhood of the test sample), the test sample is assigned to class m if $k_m = \max\{k_i, i = 1, 2, 3\}$. The algorithm used is as follows.
Algorithm 3:

Step 1) Obtain and store the distances between the test sample and all the samples in the design set.

Step 2) Sort the obtained distance values in ascending order.

Step 3) Consider the subset of the first k distances in the sorted array; i.e., the k-NN. Knowing the class membership of each of these samples, assign the test sample to the majority class in this subset if it exists, otherwise the result is considered inconclusive.

Step 4) Estimate the error rate by comparing the classification results with actual class membership. Treat the special case of inconclusive decisions individually as a separate entity (i.e., neither an error nor a correct decision) and obtain its rate of occurrence.

C) Neural Network Classifiers

In designing neural network classifiers emphasis has been on the comparison between different architectures. Learning parameters are fixed for different architectures used in this paper to neutralize their effect on the results. A bipolar neural activation function is taken to be in the form

$$f(net) = \frac{2}{1 - \exp[-\lambda net]} - 1 \quad (26)$$

where net is the sum of all inputs to the neuron multiplied by their weights and λ is the activation constant which is set to be one in all architectures. As a first preprocessing step, the training set is normalized between ± 1 for all of the parameter values. This can be performed in the same way as described in the previous section. This effectively helps speeding up the training by moving our operating point to the portion of the neuron activation function which has the highest slope. Also, a bias weight is added to each case expanding the size of the input space by one. This is important in order to provide the perceptrons with the ability to produce classification hyperplanes not necessarily passing through the origin, and to enhance the efficiency of the training algorithm. Also, each of the pathologies is assigned a binary representation with the possibility to take only two values: present/absent which are mutually exclusive. Several architectures were investigated in this research and are summarized below.

1) Single-Layer Perceptrons: In this method, a single perceptron is assigned for each pathology. This particular perceptron produces an output of one for present and −1 for absent. The algorithm used to train this class is based on delta learning rule which can be summarized as follows.

Algorithm 4:

Step 1) Arrange the training data set in the form of three separate data subsets. For each data sample, the data subset i should contain the eight parameter values augmented by an additional constant value, taken to be $+1$, as the input vector $\mathbf{x}_i$, and corresponding to one of the pathologies desired output d_i. The pairs $\mathbf{x}_i$ and d_i will be used to train the single perceptrons independently. Small initial values for the weight vectors $\mathbf{W}_i$'s are assumed for the three-perceptron layer (usually taken as random values of uniform distribution in order of magnitude of less than the inverse of the fan-in of the connected neuron).

Step 2) Apply the samples of the training sets to the inputs of the layer. Calculate their corresponding outputs $f(net_i)$.

Step 3) Obtain the diagnosis produced by the layer by either using a high-gain output stage to binarize the activation values into ± 1, and choose the diagnosis corresponding to the neuron of output $= 1$, or by selecting the maximum activation to be the diagnosed pathology. Compare to the desired value and determine whether the diagnosis is correct or not for this case.

Step 4) Calculate the updated weight vectors for each neuron as

$$\Delta \mathbf{W}_i = \frac{C}{2} [d_i - f(net_i)][1 - f^2(net_i)]\mathbf{x}_i \quad (27)$$

where C is a learning constant fixed as 0.1.

Step 5) Perform Steps 2) and 3) for all training pairs in the data sets. Evaluate the diagnosis rate for the whole data set. If all responses are correct then stop, otherwise continue network training.

Step 6) Stop the training and declare an overrun condition if the number of training cycles exceeds a certain limit chosen here to be 10 000. In this case, the network is considered to be incapable of separating the data classes, and the weight values corresponding to the best diagnostic rate during the training are selected.

Step 7) Test the network after training by classifying the samples from the test set and obtain the error rate for the network by the percentage of the number of misclassified samples to the total number of test samples.

2) Multilayer Perceptron Architectures: Multilayer perceptron network consists of a first (hidden) layer of perceptrons feeding a second (output) layer of size equal to the number of pathologies. An illustration of this type of networks is shown in Fig. 6. Various hidden layer sizes are used and the outputs are produced by binary or maximum response decision criteria to decide one of the three pathologies. The algorithm used to train this classifier is the multilayer error back-propagation algorithm which can be summarized as follows [22].
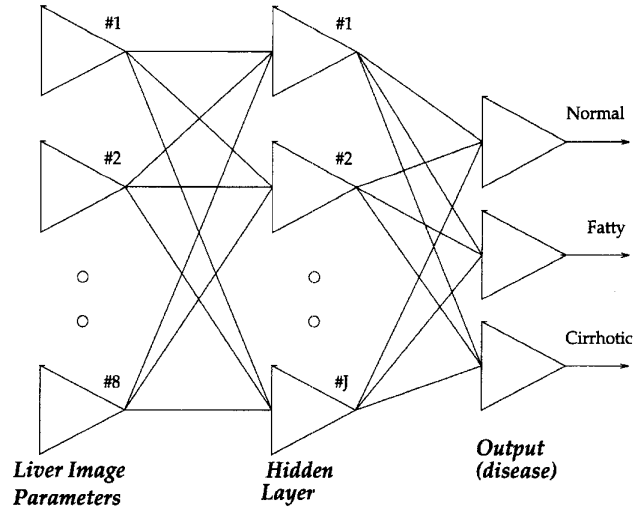

Fig. 6. Multilayer neural network structure. It has an intermediate ("hidden") layer of neurons to enhance its correct classification capabilities.
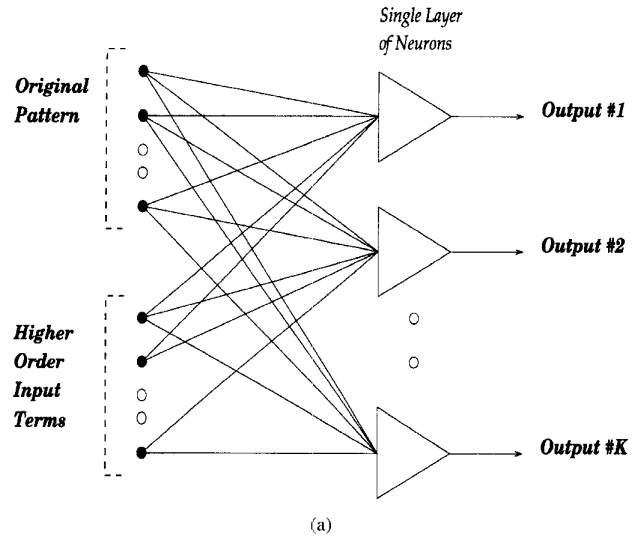

(a)

Fig. 7. Functional link network with single layer of neurons with expanded inputs including the cross product terms of the original parameters. (a) General structure.

Algorithm 5:

Step 1) Initialize the weight vectors of the two layers of the network as in Step 1), Algorithm 4.

Step 2) Introduce the training samples to the inputs of the network and calculate the outputs of the first layer. Then, these outputs after augmentation by a constant value are the inputs to the second layer. Hence, the outputs of the second layer can also be calculated.

Step 3) The weight vectors update for the second layer neurons will follow exactly the Steps 3)–5) of Algorithm 4 replacing the input vector $\mathbf{x}$ by the vector of outputs of the first layer. On the other hand, the weight update for the first layer can be expressed as follows: define the output layer error
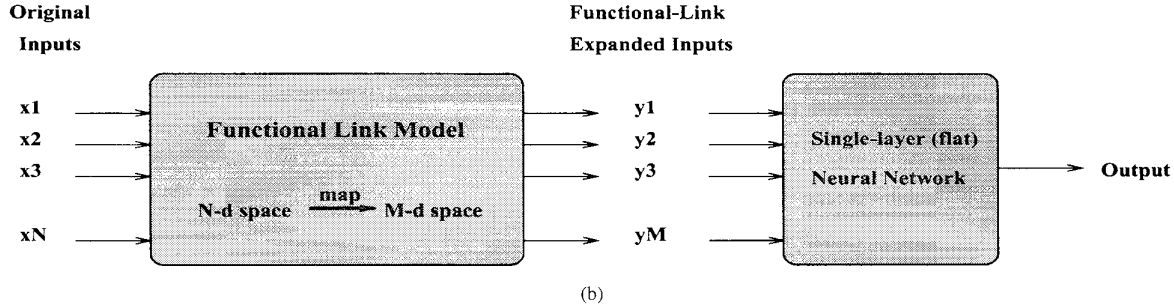
Original Inputs

Functional-Link Expanded Inputs

(b)

Fig. 7. *(Continued)*. Functional link network with single layer of neurons with expanded inputs including the cross product terms of the original parameters. (b) Network implemented in this paper.

signal

$$\varepsilon^{(2)} = \tfrac{1}{2} \{d_i - f[net_i^{(2)}]\} \{1 - f^2[net_i^{(2)}]\} \quad (28)$$

where $net_i^{(j)}$ is the *net* value at the input of the ith neuron in the jth layer, hence calculate the hidden layer error signal $\varepsilon_k^{(1)}$ as the sum of all output layer error signals multiplied by the connecting weight from the kth neuron in the hidden layer to all neurons in the output layer. Then update the weight for the ith neuron in the first layer as

$$\Delta \mathbf{W}_i^{(1)} = \frac{C}{2} \{1 - f^2[net_i^{(1)}]\}\varepsilon_i^{(1)}\mathbf{x} \quad (29)$$

where C is a learning constant fixed as 0.1. A momentum term may be added to the weight update which is equal to the previous weight update multiplied by a positive constant factor <1, chosen here as 0.5.

Step 4) Determine the condition to stop and test the network exactly as in Algorithm 4.

3) Functional Link Network: The functional link networks constitute a generalization of the usual neural networks in that the neurons actual inputs are not linear combinations of the given inputs, which in our case are the tissue parameters. For example, if the given parameters are A, B, and C, then the inputs for a regular network neurons may only include terms of the form $\alpha_{i1}A + \alpha_{i2}B + \alpha_{i3}C$, where α_{ij} are constants. On the other hand, the functional link network inputs may include quadratic terms (e.g., A^2) or even trigonometric functions [e.g., $\sin(B)$] of those parameters.

Several approaches have been tried to generate those higher order terms (e.g., [26]). These approaches can be in general classified into two main categories. The first one is to simulate the biological neural computing characteristics, where the real inputs from the dendrites cannot be physically separated while acting on a single neuron and a cross product terms naturally arise; the second, to use simple forms for the expansion terms to gain mathematical tractability. The basic mathematical theory in either case indicates that the functional-expansion model should converge to a flat-net solution if a large enough number of additional independent terms are used. Examples of such functions include tensor (outer-product) model and pure harmonic functions [27]. The superior performance of

TABLE I
MINIMUM DISTANCE CLASSIFIER

Pathology	Before Normalization	After Normalization
Specificity	29.4%	23.5%
Sensitivity for Cirrhotic	23.5%	0%
Sensitivity for Fatty	41.2%	76.5%

TABLE II
BAYES QUADRATIC CLASSIFIER

Pathology	Before Normalization	After Normalization
Specificity	86.5%	56.8%
Sensitivity for Cirrhotic	16.2%	0%
Sensitivity for Fatty	70.3%	91.9%

functional link networks can be demonstrated in the two-dimensional (2-D) case by the implementation of a two-input logical XOR function using binary neurons, where a single cross-term is able to realize the function which is otherwise impossible using the standard inputs. For the general case of n-dimensional input space, we can imagine that the functional expansion creates more flexible classification surfaces than the set of intersecting hyperplanes created by the linear inputs.

In our case, the two approaches of generating higher-order terms converge to the same path since we are using the tensor (outer product) model. This includes second- or third-order cross-terms to expand the input space. This model can be formally defined as follows.

a) Tensor Model A (second-order model)

$$\{x_i, \quad i = 1, 2, \cdots, 8\} \xrightarrow{\text{Model A}}$$
$$\{x_ix_j, \quad i, j = 1, 2, \cdots, 8, \quad i \neq j\}. \quad (30)$$

b) Tensor Model B (third-order model)

$$\{x_i, \quad i = 1, 2, \cdots, 8\} \xrightarrow{\text{Model B}}$$
$$\{x_ix_jx_k, \quad i, j, k = 1, 2, \cdots, 8, \quad i \neq j \neq k\}. \quad (31)$$

So, for our case, the number of added terms will be 28, which is calculated from the number of possible distinct pairs drawn from the available eight different parameters. In general, for the case of N distinct parameters, the number of added terms in the tensor model A will be $N(N - 1)/2$. Similarly for the tensor model B, the number of added terms will be $N(N-1)(N-2)/3 \times 2$. Hence, for our case, the tensor model B will add 56 higher-order terms, and the combined model

TABLE III
VOTING k-NN CLASSIFIER

k	Before Normalization			After Normalization		
	Specificity	Sens.Cirrhotic	Sens.Fatty	Specificity	Sens.Cirrhotic	Sens.Fatty
1	94.1%	100%	94.1%	88.2%	100%	100%
2	100%	100%	93.3%	88.9%	100%	100%
3	81.3%	82.4%	94.1%	62.5%	94.1%	93.8%
4	84.6%	100%	94.1%	63.6%	93.3%	93.8%
5	82.4%	94.1%	94.1%	53.3%	87.5%	93.8%
6	75.0%	92.3%	94.1%	53.8%	100%	93.8%
7	80.0%	88.2%	100%	53.3%	87.5%	88.2%
8	78.6%	88.2%	94.1%	53.8%	86.7%	88.2%
9	81.8%	88.2%	94.1%	60.0%	93.3%	88.2%

TABLE IV
VOTING k-NN CLASSIFIER INCONCLUSIVE DECISION RATES

k	Before Normalization			After Normalization		
	Normal	Cirrhotic	Fatty	Normal	Cirrhotic	Fatty
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	47.0%	29.4%	11.8%	47.0%	5.9%	11.8%
3	5.9%	0.0%	0.0%	5.9%	0.0%	5.9%
4	23.5%	17.6%	0.0%	35.3%	11.8%	5.9%
5	0.0%	0.0%	0.0%	11.8%	5.9%	5.9%
6	29.4%	23.5%	0.0%	23.5%	29.4%	5.9%
7	11.8%	0.0%	5.9%	11.8%	5.9%	0.0%
8	17.7%	0.0%	0.0%	23.5%	11.8%	0.0%
9	35.3%	0.0%	0.0%	11.8%	11.8%	0.0%

TABLE V
SINGLE PERCEPTRON TRAINING

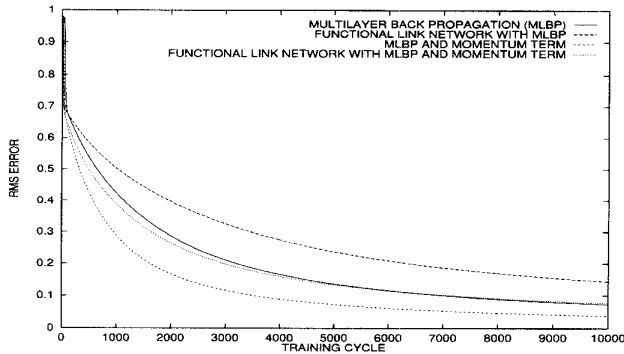| Pathology	Training Set	Test Set
Specificity	86.7%	88.3%
Sensitivity for Cirrhotic	95.0%	91.7%
Sensitivity for Fatty	100%	96.7%

Fig. 8. Training profile of multilayer neural network with five hidden-layer neurons. The momentum term has a positive effect to speed up the training while the functional link inputs seems to have a rather negative effect.
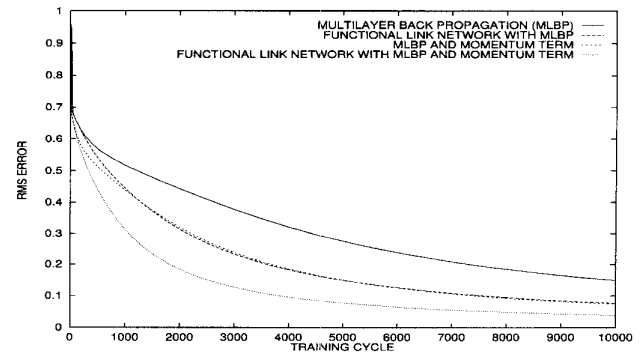

Fig. 9. Training profile of multilayer neural network with 20 hidden-layer neurons. Here, both the momentum term and the functional link inputs seem to improve the speed of the network training.
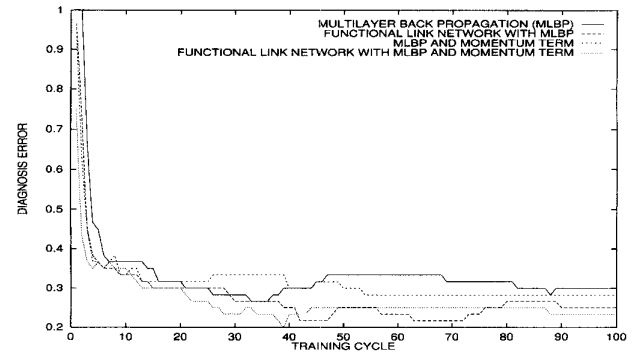

Fig. 10. Training profile of multilayer neural network with 20 hidden-layer neurons. The effect of adding the functional link inputs is shown to improve the diagnosis error more than adding the momentum term alone.

(A) and (B) will add 84 terms. A graphical illustration of the general idea of functional link inputs is shown in Fig. 7(a). Also, the architecture of the network used in our application is shown in Fig. 7(b).

4) Data Compression Using Cluster Center of Gravity: In this Section, we compress our data set to only one record per pathology, which is the center of gravity (CG) of the pathology cluster, in order to obtain easier training. The center of gravity of a particular cluster is defined as the weighted mean of all n-dimensional parameter vectors belonging to the set of all cases with a particular pathology forming the cluster. It is important to note that the diagnosis errors obtained are not of the same weight like we usually have in other situations. As indicated before, liver diseases are of a serious nature and can be life threatening. Therefore, the misclassification errors are categorized into false negative, which means that a patient is diagnosed as normal while the patient has a disease, and false positive, which means that a normal patient is diagnosed to have a disease. Obviously, the first category is most dangerous and should ultimately be eliminated. Consequently, this can be an important point for correct interpretation of the results.
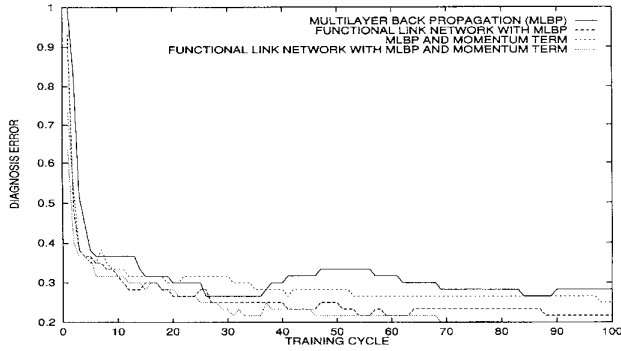
Fig. 11.   Training profile of multilayer neural network with 50 hidden-layer neurons. Diagnosis error improves with increasing the size of the hidden-layer.

Fig. 13.   Training profile of a three-neuron, single layer, functional link neural network tensor models (A) and (B). Training speed improves for the cirrhotic and fatty cases but the training for normal cases is slightly degraded.

TABLE VI
MULTILAYER NETWORK TRAINING—SUCCESS
RATES USING 1000 TRAINING CYCLES

Training Algorithm	Hidden-Layer Size	Training Set	Test Set
MLBP*	5	73.3%	68.3%
MLBP+Momentum Term	5	75.0%	68.3%
MLBP	20	76.7%	68.3%
MLBP+Momentum Term	20	73.3%	70.0%

* (Multilayer Back Propagation Algorithm)

TABLE VII
FUNCTIONAL LINK NETWORK—TENSOR MODEL (A)

Pathology	Training Set	Test Set
Specificity	100%	90.0%
Sensitivity for Cirrhotic	100%	96.7%
Sensitivity for Fatty	100%	96.7%

TABLE VIII
FUNCTIONAL LINK NETWORK—TENSOR MODEL (A) & (B)

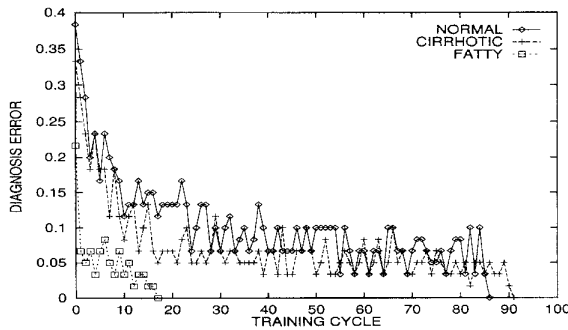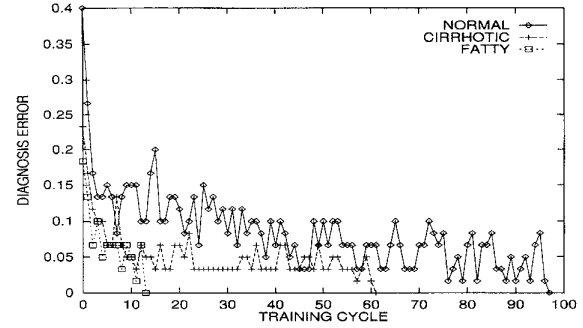| Pathology	Training Set	Test Set
Specificity	100%	93.3%
Sensitivity for Cirrhotic	100%	98.4%
Sensitivity for Fatty	100%	100%

Fig. 12.   Training profile of a three-neuron, single layer, functional link neural network tensor model (A). Diagnosis error reaches zero after 90 training cycles only for the three pathologies.

IV. RESULTS

A. Statistical Classifiers

1) Minimum Distance Classifier: The results of applying this classifier are not satisfactory at all. The correct diagnosis rates for different pathologies are listed in Table I. Obviously, this classifier is not suitable for this application. This can be explained by the presence of a high degree of overlapping among the three pathology data clusters, which simply could not be resolved using only hyperplanes as assumed by this classifier.

2) Bayes Quadratic Classifier: The results for this classifier are listed in Table II and are better than that of type A. However, the rate of correct diagnosis is still not sufficient for practical use. This could be related to both insufficient accuracy of estimation due to the low sample volume and to improper initial assumptions of probability distributions.

3) Voting k-NN Classifier: The results of this method are shown in Table III. As expected, this method yields much better results than both former methods to classify the three pathologies. This can be explained by the inherited independence of the technique from the data distribution by being sample-based. The rate of inconclusive decisions, shown in Table IV, should be carefully considered because it indicates how often this classifier becomes useless. In other words, the rate of inconclusive decisions should be as small as possible for a reliable classifier. Therefore, the best performance is obviously obtained for $k = 1$, sometimes called nearest neighbor classifier.

B. Neural Network Classifiers

1) Single-Layer Perceptron Network: This architecture showed a relatively poor response. The results obtained are shown in Table V. It can be seen that the error is 22% with false-positive rate of half this number, which is quite unacceptable for good diagnosis. This performance is mainly due to the high degree of overlapping between the three pathology clusters, which goes beyond the linear separability provided by the single-layer perceptrons [22].

2) Multilayer Perceptron Network: This architecture showed high performance for steady error reduction during training. However, the network requires much longer training time due to the large number of input patterns (60 different patterns), in addition to the large size of weight matrices, especially

TABLE IX
SUCCESS RATE USING CG'S ONLY TO TRAIN MULTILAYER NEURAL NETWORKS

Training Algorithm	Hidden layer	Error for Training Set	Error for Test Set
MLBP	5	100%	61.7%
MLBP+Momentum Term	5	100%	61.7%
MLBP	20	100%	60.0%
MLBP+Momentum Term	20	100%	65.0%

for a large number of hidden-layer neurons. The introduction of a momentum term into the training has slowed down the training even more while reducing the error. In addition, the network training may be diversified even more to improve the convergence of the algorithm and to avoid local minima. This can be achieved by using simulated annealing technique to find the global minimum error. Apart from these problems, the performance of this type of networks is good for binary output decision layer and 30% better for maximum-response decision layer. Figs. 8–11 illustrate various examples of training profiles of this category. The results after 1000 training cycles are shown in Table VI. As can be seen, the introduction of the momentum term helps to reduce the RMS error more quickly. However, this architecture is not the best one for the class of problems because the training has been found to be rather lengthy.

3) Single-Layer Perceptron Network with Functional Link Inputs: Using the additional tensor model inputs (A) and (B) to the normal input space, excellent results are obtained even with a single layer of perceptrons. The training profile is shown in Figs. 12 and 13, and the results are listed in Tables VII and VIII. Results of this classifier are obviously quite good; they are based on a simple network with only a single layer and with full ability to add a new pathology independent from the other existing pathologies. The explanation for superior performance of this network is that the introduction of new terms enables the neurons to have nonlinear classification capabilities, which are originally not present in neural network architectures with single layer. This is achieved through adding the product terms $x_i x_j$ in the *net* function, which represents hyperbolic decision lines in all planes parallel to the $x_i x_j$ plane.

4) Multilayer Perceptron Network with Functional Link Inputs: Training results of this classifier are illustrated in Figs. 8–11 in comparison with other classifiers. It is shown that the introduction of additional terms to a multilayer network with only few perceptrons in the hidden layer degrades the performance of the training. On the other hand, these terms may be of significance for multilayer networks having hidden layers of large dimension.

5) Validity of Using Cluster Center of Gravity for Training: The cluster CG's are obtained for the three pathology clusters and used as the only training set for the multilayer perceptron network. The main purpose of this experiment is to try to find out a method to overcome the slow rate of learning for this network. The results of the experiment are summarized in Table IX. Obviously, the zero error for the training set does not imply that the network can work as well for data around the training CG's. Hence, this method did not provide acceptable results for this task.

V. CONCLUSIONS

The primary conclusion of this research is to confirm the importance of the choice of classification techniques on the success rate of QTCT. Some of the sound classification techniques (e.g., the Bayes minimum error classifier) failed to provide acceptable results, while others could. This might also be a significant indication that the design of classifiers assuming normality of data alone can lead to poor results due to the complexity and overlapping between different pathology clusters. Results of this research showed a good promise for functional link neural networks in automatic diagnosis. Although the method does not actually expand the input space (because all additional terms can be proven to be dependent), it has a powerful capability to create hyperbolic surfaces in addition to the original straight lines as decision boundaries. Also, the high performance of the sample-based voting k-NN is appreciable. We believe that these two statistical and neural methods offer high degree of efficiency due to their simple implementation and use as compared to other more complex techniques. Normalizing the classification data has been shown to improve the overall diagnostic accuracy, indicating its importance as a preprocessing step. Further progress can be achieved in the functional link approach if one tries to minimize the number of added terms to achieve the required correct classification. This may be done through elimination of the elements multiplied by the smallest weights and then repeating the training procedure again. Also, the effect of perceptron activation and training parameters may further be studied to obtain the set of parameters yielding the best performance for a given architecture. On the other hand, in the k-NN method, we can resolve the problem of inconclusive decisions by assigning a certain weight to each of the k neighbors that is inversely proportional to their distances to the sample of interest. These weights are used to favor certain decision in the absence of a majority class. The findings of this paper may also suggest a much easier hardware implementation of tissue analysis functions to be provided in ultrasound machines in the future.

REFERENCES

[1] D. A. Christensen, *Ultrasound Bioinstrumentation*. New York: Wiley, 1988.
[2] A. M. Youssef and A. A. Sharawi, "KSODATA clustering analysis for diffuse liver diseases," in *Proc. IEEE Symp. Ultrasound*, 1990.
[3] D. Schlaps *et al.*, "Ultrasonic tissue characterization using a diagnostic expert system," in *Information Processing in Medical Imaging*, S. L. Bachrach, Ed. Dordecht, Germany: Martinus Nijhoff, 1986.
[4] B. S. Garra, M. F. Insana, T. H. Shawker, and M. A. Russell, "Quantitative estimation of attenuation and echogenicity," *Radiol.*, pp. 1661–1667, 1987.

[5] K. J. Parker, "Ultrasonic attenuation and absorption in liver tissue," *Ultrasound Med. Biol.,* vol. 9, pp. 363–369, 1983.

[6] C. B. Burchardt, "Speckle in ultrasound B-mode scans," *IEEE Trans. Sonics Ultrason.,* vol. SU-25, pp. 1–6, 1978.

[7] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans," *IEEE Trans. Sonics Ultrason.,* vol. SU-30, pp. 156–163, 1983.

[8] M. F. Insana, B. S. Garra, D. G. Brown, and T. S. Shawker, "Analysis of ultrasound images via generalized Rician statistics," *Opt. Eng.,* vol. 25, pp. 743–748, 1986.

[9] E. Walach, A. Shmulewitz, Y. Itzchak, and Z. Heyman, "Local tissue attenuation images based on pulsed-echo ultrasound scans," *IEEE Trans. Biomed. Eng.,* vol. 36, no. 2, pp. 211–220, 1989.

[10] E. Walach, C. N. Liu, R. C. Waag, and K. J. Parker, "Quantitative tissue characterization based on pulse-echo-ultrasound scanning," *IEEE Trans. Biomed. Eng.,* vol. BME-33, pp. 637–644, 1986.

[11] S. Fields and F. Dunn, "Correlation of echographic visualizability of tissue with biological composition," *J. Acoust. Soc. Amer.—Med. pt. I: Basic principles, special and biological state,* vol. 54, pp. 809–812.

[12] D. Nicholas, *In Recent Advances in Ultrasound in Biomedicine,* D. N. White, Ed. Oregon: Research Studies, ch. 1, pp. 1–28.

[13] A. M. Youssef and Y. M. Kadah, "Ultrasonic tissue characterization of breast masses using a diagnostic expert system," in *Proc. Comput. Assist. Radiol., CAR-89,* Berlin, Germany, June 1989.

[14] A. M. Youssef, D. Schlaps, and W. J. Lorenz, "Ultrasound textural synthesis using 2-D autoregressive models for pathology characterization," in *Proc. SPIE-87,* vol. 768, 1987, pp. 201–206.

[15] B. S. Garra, *In Vivo Liver and Splenic Tissue Characterization by Scattering,* book in preparation.

[16] B. B. Gosnik, S. K. Lemon, W. Scheible, and G. R. Leupold, "Accuracy of ultrasonography in diagnosis of hepatocellular disease," *AJR,* vol. 133, pp. 19–23, 1979.

[17] K. J. Foster, K. C. Dewbury, A. H. Griffith, and R. Wright, "The accuracy of ultrasound in the detection of fatty infiltration of the liver," *Br. J. Radiol.,* vol. 53, pp. 440–442, 1980.

[18] A. M. Badawi and A. M. Youssef, "Effect of static compression on textural and acoustical parameters of ultrasound images," in *Symp. Egyptian Soc. Hepatol.,* 1992.

[19] A. Sharawi, "Quantitative tissue characterization parameters for liver diseases," Ph.D. thesis, Systems and Biomed. Eng. Dept, Faculty of Eng., Cairo Univ., 1990.

[20] Y. M. Kadah, A. A. Farag, A. M. Youssef, and A. S. Badawi, "Automatic tissue characterization from ultrasound imagery," in *Proc. SPIE-93,* Boston, MA, 1993, vol. 2055.

[21] ———, "Statistical and neural classifiers for ultrasound tissue characterization," in *Proc. ANNIE-93, Artificial Neural Networks in Engineering,* Rolla, MO, Nov. 1993. Also in *Intelligent Engineering Systems Through Artificial Neural Networks,* vol. 3, C. H. Dagli *et al.,* Eds. New York: ASME, 1993, pp. 397–402.

[22] J. M. Zurada, *Introduction to Artificial Neural Systems.* Boston, MA: PWS, 1992.

[23] H. L. van Trees, *Detection, Estimation, and Modulation Theory, Pt. I.* New York: Wiley, 1968.

[24] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2nd ed. New York: Academic, 1990.

[25] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* New York: Wiley, 1973.

[26] Y. Pao, *Adaptive Pattern Recognition and Neural Networks.* New York: Addison-Wesley, 1989.

[27] Y. Pao and Y. Takefuji, "Functional-link net computing: Theory, system architecture, and functionalities," *Comput.,* vol. 25, no. 5, pp. 76–79, May 1992.

[28] T. Zielke, P. Nauth, N. Stein, W. von Seelen, E. G. Loch, and A. Gaca, "Quantitative techniques in ultrasonic diagnosis," *Radiologe,* vol. 25, no. 10, pp. 486–73, 1985.

[29] E. Walach, C. N. Liu, R. C. Waag, and K. J. Parker, "Quantitative tissue characterization based on pulsed-echo ultrasound scans," *IEEE Trans. Biomed. Eng.,* vol. BME-33, no. 7, pp. 637–643, 1986.

[30] D. Nicholas, D. K. Nassiri, P. Garbutt, and C. R. Hill, "Tissue characterization from ultrasound B-scan data," *Ultrasound Med. Biol.,* vol. 12, no. 2, pp. 135–143, 1986.

[31] M. J. Muller, D. Lorenz, I. Zuna, W. J. Lorenz, and G. van Kaick, "The value of computer-assisted sonographic tissue characterization in focal lesions of the thyroid," *Radiologe,* vol. 29, no. 3, pp. 132–136, 1989.

[32] U. Raeth, D. Schlaps, B. Limberg, I. Zuna, A. Lorenz, G. van Kaick, W. J. Lorenz, and B. Kommerell, "Diagnostic accuracy of computerized B-scan texture analysis and conventional ultrasonography in diffused parenchynal and malignant liver disease," *J. Clinical Ultrasound,* vol. 13, no. 2, pp. 87–99, 1985.

[33] M. F. Insana, B. S. Garra, S. J. Rosenthal, and T. J. Hall, "Quantitative ultrasonography," *Med. Progress Through Technol.,* vol. 15, no. 3/4, pp. 141–153, 1989.

[34] B. S. Garra, M. F. Insana, T. H. Shawker, R. F. Wagner, M. Bradford, and M. Russel, "Quantitative ultrasonic detection and classification of diffuse liver disease. Comparison with human observer performance," *Investigative Radiol.,* vol. 24, no. 3, pp. 196–203, 1989.

[35] U. Haberkorn, I. Zuna, A. Lorenz, H. Zerban, G. Layer, G. van Kaick, and U. Rath, "Echographic tissue characterization in diffuse parenchymal liver disease: Correlation of image structure with histology," *Ultrason. Imag.,* vol. 12, no. 3, pp. 155–170, 1990.

[36] T. A. Tuthill, R. B. Baggs, and K. J. Parker, "Liver glycogen and water storage: Effect on ultrasound attenuation," *Ultrasound Med. Biol.,* vol. 15, no. 7, pp. 621–627, 1989.

[37] K. L. Kader, S. Hubbard, and B. A. Carroll, "Hepatic ultrasonic tissue characterization of video image data," *Investigative Radiol.,* vol. 25, no. 10, pp. 1091–1094, 1990.

[38] E. J. Halpern, "Effect of focal zone on a curved array transducer," *Investigative Radiol.,* vol. 28, no. 1, pp. 7–10, 1993.

[39] U. Haberkorn, G. Layer, V. Rudat, I. Zuna, A. Lorenz, and G. van Kaick, "Ultrasound image properties influenced by abdominal wall thickness and composition," *J. Clin. Ultrasound,* vol. 21, no. 7, pp. 423–429, 1993.