Research Article

Linear Classifier with Reject Option for the Detection of Vocal Fold Paralysis and Vocal Fold Edema

Constantine Kotropoulos (EURASIP Member)[1, 2] and Gonzalo R. Arce[2]

[1] *Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki 54124, Box 451, Greece*
[2] *Department of Electrical and Computer Engineering, University of Delaware, 140 Evans Hall, Newark, DE 19716, USA*

Correspondence should be addressed to Constantine Kotropoulos, costas@zeus.csd.auth.gr

Two distinct two-class pattern recognition problems are studied, namely, the detection of male subjects who are diagnosed with vocal fold paralysis against male subjects who are diagnosed as normal and the detection of female subjects who are suffering from vocal fold edema against female subjects who do not suffer from any voice pathology. To do so, utterances of the sustained vowel "ah" are employed from the Massachusetts Eye and Ear Infirmary database of disordered speech. Linear prediction coefficients extracted from the aforementioned utterances are used as features. The receiver operating characteristic curve of the linear classifier, that stems from the Bayes classifier when Gaussian class conditional probability density functions with equal covariance matrices are assumed, is derived. The optimal operating point of the linear classifier is specified with and without reject option. First results using utterances of the "rainbow passage" are also reported for completeness. The reject option is shown to yield statistically significant improvements in the accuracy of detecting the voice pathologies under study.

1. Introduction

Vocal pathologies arise due to accident, disease, misuse of the voice, or surgery affecting the vocal folds and have a profound impact on patients' life. The modeling of normal and pathological voice source and the analysis of healthy and pathological voices has gained increasing interest recently [1]. Among the most interesting works are those concerned with Parkinson's Disease (PD) and multiple sclerosis, which belong to a class of neurodegenerative diseases that affect patients speech, motor, and cognitive capabilities [2, 3]. People with neurological conditions causing disability often have associated dysarthria, which is the most common acquired speech disorder affecting 170 per 100 000 population [4]. Several studies explore the main voice characteristics (i.e., the fundamental frequency and vocal tract resonance frequencies) together with their deviation from the nominal conditions for persons who exhibit voice disorders. Although the majority of techniques analyze the speech signal, the video modality offers complementary information [5, 6]. For example, three-dimensional (3D) magnetic resonance imaging could be used to build a 3D numerical model of the vocal tract and videokymography could overcome the transmission speed and volume limitations of 2D imaging (i.e., stroboscopy) for severely dysphonic patients with an aperiodic signal, allowing to register the movements of the vocal folds with a high time resolution on a line perpendicular to the glottis [1]. Furthermore, the irregular vocal fold oscillations can be observed by means of a digital high-speed camera using image processing techniques in order to extract the vocal fold edges, estimate the minimum glottal area defined by the vocal fold positions, and compute the distance between the glottal midline and the vocal fold edges extracted at medial position in real-time [7]. The time series of such displacements can drive an inversion procedure in order to adjust the parameters of a biomechanical model of vocal folds for both pathological and healthy vocal fold oscillations. All the aforementioned techniques aim at evaluating the performance of special treatments, such as the Lee Silverman Voice Treatment [3], assisting the e-inclusion of people with physical disabilities and disordered speech by offering better access to telecommunication services [8] or

more efficient environmental control systems [9]. Thus, it is a matter of great significance to develop systems able to classify the incoming voice samples as normal or pathological ones before other procedures are further applied.

Voice pathologies may be assessed by either perceptual judgments or an objective assessment. The perceptual judgment resorts to qualifying and quantifying the vocal pathology by listening to patients' speech. Although this is the most commonly used method by clinicians, it suffers from several drawbacks. First of all, the perceptual judgment has to be performed by an expert jury in order to increase its reliability. Second, due to the lack of universal assessment scales and the dependence on experts' professional background and experience or the knowledge of patients history, the perceptual judgment may involve large intra and inter-variability. Third, the perceptual analysis is very costly in time and human resources and cannot be planned regularly. Nowadays an increasing use of objective measurement-based analysis as a non-invasive technique for supporting diagnosis in laryngeal pathology has been observed [8–11]. Objective measurement-based analysis qualifies and quantifies the voice pathology by analyzing acoustical, aerodynamic, and physiological measurements. These measurements may be directly extracted from patient's speech utterance using a simple computer-based system or may require special instruments. Typical techniques, such as fundamental frequency and jitter estimation should be carefully adapted in order to take into account the significant variations of fundamental frequency from cycle to cycle as well as the presence of subharmonic and aperiodic components in the pathological voice [12–14]. Very useful insight to the production of disordered speech could be obtained through simulation studies [15–17]. Although the objective analysis alleviates the subjectivity of perceptual judgments, it has certain limitations as well. First, the objective analysis often relies on pattern recognition techniques, such as linear discriminant analysis, correlation estimation, which do depend on the measurements being analyzed. Second, the objective analysis is frequently confined to the study of sustained vowels only, which are not representative of continuous speech [18]. In the medical literature, agreement between the perceptual judgments and the findings of objective analysis is generally sought for [19, 20].

Several techniques for the detection and classification of voice pathologies by means of acoustic analysis, parametric and non-parametric feature extraction, and pattern recognition are reviewed in [21]. In all these techniques, first, descriptive features are extracted from the speech signal. A number of so-called classical parameters quantify pitch perturbations (jitter), amplitude perturbations (shimmer) and estimate the Harmonic to Noise Ratio at different frequency bands and the critical-band energy spectrum by employing either short-term Discrete Fourier Transform and cepstral analysis [22–24] or the singularities in the power spectral density of the vocal cord cover wave (also referred to as the mucosal wave correlate) [25]. Alternatively, features stemming from the 1-D bicoherence index derived by the bispectrum [22] or nonlinear dynamical system theory, such as statistics of the correlation dimension and the largest Lyapunov exponent [26], or the return period density entropy [27] were extracted. Features could also be obtained by applying the continuous wavelet transform to each speech frame and averaging neighbor wavelet coefficients on time-frequency scale [28]. Frequently, feature vectors undergo dimensionality reduction by applying Principal Component Analysis (PCA) [29–31] before classification or a subset of features are selected by applying either a wrapper or a filter. Next, the features are either clustered in a number of pre-defined classes, say by a K-means algorithm [30] or are fed to a classifier, which is designed to solve a two-class pattern recognition problem. That is, to verify a specific pathology in a test utterance or to decide whether a test utterance is pathological or not. Commonly used classifiers resort to linear discriminant analysis (LDA) [23, 27, 29, 32], nearest neighbors [24, 26, 29], vector quantization [33] or support vector machines (SVMs) [28, 31, 34]. It is worth noting that the detection of voice pathology is closely related to speaker verification. In particular, pathological class models can be derived from generic Gaussian mixture models by employing the maximum a posteriori adaptation technique [35] and adapting only the means [34]. While a sustained phonation can be classified as normal or pathological with an accuracy greater than 90% when speech is recorded in laboratory conditions [21], telephone quality speech can be classified as normal or pathological with a much smaller accuracy, that is, 74.15% [23].

In this paper, we are concerned with vocal fold paralysis and vocal fold edema, which are both associated with communication deficits that affect the perceptual characteristics of pitch, loudness, quality, intonation, and have similar symptoms with PD and other neuro-degenerative diseases [36]. We are interested in detecting male subjects who are diagnosed with vocal fold paralysis against male subjects who are diagnosed as normal. Similarly, we would like to distinguish between female subjects who are diagnosed with vocal fold edema against female subjects who are diagnosed as normal. Utterances from the Massachusetts Eye & Ear Infirmary (MEEI) Voice Disorders Database, which is distributed by Kay Elemetrics [37], are employed, because the MEEI database is a benchmark annotated speech corpus. A review of several voice pathology detection approaches with the MEEI database can be found in [21]. However, the majority of these approaches aim at identifying whether an utterance is pathological or not without addressing which speech pathology is observed. Although a direct comparison between these methods is not possible, because different data subsets have been used and different performance criteria have been employed, one can roughly claim that the state of the art accuracy in detecting whether an utterance is pathological or not exceeds 98% [38, 39]. In the following, let us confine ourselves to vocal fold paralysis and edema detection. The identification of vocal fold paralysis using the normalized energy across various scaling factors of the wavelet transform and a multilayer neural network trained by back-propagation was proposed [40]. For 50 data samples of the MEEI database, an average classification accuracy of 90% was reported. The performance of Fisher's linear classifier, the K-nearest neighbor classifier, and the

nearest mean one for detecting vocal fold paralysis in male utterances and vocal fold edema in female utterances was assessed in [29]. The subjects were called to articulate the sustained vowel "ah" (/a/). From each recording, two central frames were selected among the ones that belong to the most stationary portion of the sustained speech signal as is proposed in [41, 42]. 14-order linear prediction coefficients (LPCs) were extracted from each frame. The dimensionality of the raw feature vector was then reduced to 2 by PCA. Receiver operating characteristic (ROC) curves for the Fisher linear classifier were demonstrated. It was shown that a probability of detection close to 85% could be achieved for a probability of false alarm 10% in the case of vocal fold paralysis in male utterances, while the probability of detection for vocal fold edema in female utterances was found to be approximately 73% at the same probability of false alarm. The nearest mean classifier was found to outperform K-nearest neighbor classifiers for $K = 1, 2, 3$ in both experiments. Two linear classifiers were examined in [32]. The first one is based on a sample-based optimal linear classifier design [43], while the second one is based on the dual-space linear discriminant analysis [44]. Again 14 LPCs were extracted by processing utterances corresponding to the sustained vowel "ah." Both the rectangular and the Hamming window are used to extract the speech frames [45]. The assessment of the classifiers studied in [32] was done by estimating the probability of false alarm and the probability of detection using the leave-one-out method. The parametric classifier was found to be more accurate than the dual space linear discriminant classifier. In particular, a slightly higher probability of detection for vocal fold paralysis in men was measured, that is approximately equal to 90% for probability of false alarm 10%. The gain in the probability of detection for vocal fold edema in women was 20% higher than that achieved by the Fisher linear discriminant in [29]. LPCs, LPC-derived cepstral coefficients, and mel frequency cepstal coefficients were extracted for vocal fold edema detection in [33]. A vector quantizer was trained based on the distance between the feature vectors. Experiments were conducted by using 53 normal speakers and another 67, who were diagnosed with voice pathologies including vocal fold edema. Only a single operating point was reported, which yields probability of detection approximately 73% for probability of false alarm 4% [33]. For the same probability of false alarm, a probability of detection, which falls between 80.95% for rectangular window and 90.47% for Hamming window, was reported in [32].

Two distinct two-class pattern recognition problems are studied, namely, the detection of male subjects who are diagnosed with vocal fold paralysis against male subjects who are diagnosed as normal and the detection of female subjects who are suffering from vocal fold edema against female subjects who do not suffer from any voice pathology. The rationale for gender-dependent voice pathology detection is in the inherent differences of the speech production system for male and female speakers and the higher accuracy for speech emotion recognition, speaker indexing, speaker recognition, and so forth, offered by the gender-dependent models than the gender-independent ones. The ROC curve

of the linear classifier, that stems from the Bayes classifier when Gaussian class conditional probability density functions with equal covariance matrices are assumed, is derived. The optimal operating point of the linear classifier is specified with and without reject option. The contribution of this paper is in the assessment of the impact of reject option in the ROC curve of the linear classifier for the two-class pattern recognition problems under study. Although sustained vowels are not representative of continuous speech, utterances of the sustained vowel "ah" from the MEEI database are employed here due to their wide use in medical practice and, primarily, in order to maintain direct compatibility with previously reported results [29, 32] and minimal problem complexity, so that we focus on the role of the reject option. However, first experimental results using continuous speech utterances are reported for completeness. A reject region in classifier design was also proposed in [27], but without demonstrating its impact in the ROC curve. The motivation behind the introduction of reject option in classifier design is two-fold: First, when the conditional error given a feature vector due to the decision rule (also known as classification risk) is high, the classifier should postpone making any decision and request rather for expert's advice. Second, new classes may appear during the test phase, which were not present during training or some classes may be sampled poorly during training leading to inaccurate class models [46]. The introduction of reject option in the design of two-class classifiers (also known as dichotomizers) and its impact on the ROC has recently attracted the attention of the pattern recognition community [46–49]. Linear prediction coefficients extracted from the utterances are used as features. The reject option is shown to yield statistically significant improvements in the accuracy of detecting the voice pathologies under study.

The outline of the paper is as follows. Section 2 describes briefly the Bayes classifier for both minimum error and minimum cost classification in a two-class pattern recognition problem without a reject option and discusses the motivation behind the adoption of a linear classifier. Section 2.1 defines the ROC curve and its use to derive the optimal operating point for a two-class classifier. The introduction of reject option in a dichotomizer is addressed in Section 3. The data-set used is presented in Section 4 along with feature extraction. Experimental results are reported in Section 5 and conclusions are drawn in Section 6.

2. The Bayes and the Linear Classifiers without Reject Option

Let X denote a sample (i.e., a feature vector). Let the class Ω_1 comprise of samples from healthy subjects and the class Ω_2 comprise of samples from subjects diagnosed with certain pathologies. The *Bayes rule for minimum error* assigns X to the class Ω_i having the maximum a posteriori probability given X [43]. That is,

$$\ell(X) = \frac{p_1(X)}{p_2(X)} \underset{\Omega_2}{\overset{\Omega_1}{\gtrless}} \frac{P_2}{P_1}, \tag{1}$$

where $p_i(X)$ are the class conditional probability density functions (pdfs) and P_i are the a priori probabilities of the classes Ω_i, $i = 1, 2$. The term $\ell(X)$ at the left-hand side of (1) is known as likelihood and the fraction in the right-hand side of (1) is called the threshold value of the likelihood ratio for decision [43]. Frequently, the decision is expressed in terms of the minus log-likelihood ratio $h(X) = -\ln \ell(X)$, which is known as the *discriminant function*. Let us assume that the class conditional pdfs are normal densities with mean vectors M_i and covariance matrices Σ_i, $i = 1, 2$. Then, the discriminant function becomes a quadratic function of X, that is,

$$
\begin{aligned}
h(X) = &\frac{1}{2}(X - M_1)^T \Sigma_1^{-1}(X - M_1) \\
&- \frac{1}{2}(X - M_2)^T \Sigma_2^{-1}(X - M_2) \\
&+ \frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} \underset{\Omega_2}{\overset{\Omega_1}{\lessgtr}} \ln \frac{P_1}{P_2}.
\end{aligned}
\tag{2}
$$

The minimization of the probability of classification error treats equally the misclassifications of Ω_1- and Ω_2-samples. However, a higher decision cost should be assigned whenever a patient is misclassified as normal than whenever a normal subject is misclassified as patient. By introducing the cost c_{ij} of deciding $X \in \Omega_i$ although X actually belongs to Ω_j according to ground truth, the *Bayes test for minimum cost* is obtained:

$$
\frac{p_1(X)}{p_2(X)} \underset{\Omega_2}{\overset{\Omega_1}{\gtrless}} \frac{(c_{12} - c_{22})P_2}{(c_{21} - c_{11})P_1}.
\tag{3}
$$

The comparison of (3) with (1) reveals that only the threshold has been changed in the right-hand side of the likelihood ratio test. Clearly, for symmetrical cost function, that is, $c_{12} - c_{22} = c_{21} - c_{11}$, the aforementioned likelihood ratio tests coincide. Hereafter, we will employ a *linear classifier* that stems from the quadratic one (2) if equal covariance matrices $\Sigma_1 = \Sigma_2 = \hat{\Sigma}$ are assumed, that is,

$$
\begin{aligned}
\hat{h}(X) = &\left(\widehat{M}_2 - \widehat{M}_1\right)^T \hat{\Sigma}^{-1} X \\
&+ \frac{1}{2}\left(\widehat{M}_1^T \hat{\Sigma}^{-1} \widehat{M}_1 - \widehat{M}_2^T \hat{\Sigma}^{-1} \widehat{M}_2\right) \underset{\Omega_2}{\overset{\Omega_1}{\lessgtr}} t,
\end{aligned}
\tag{4}
$$

where $\widehat{M}_i$ is the sample mean for Ω_i, $i = 1, 2$, t denotes the threshold admitting a value in the range of the discriminant function, and $\hat{\Sigma}$ is the gross sample covariance matrix estimated from the design set without making any distinction between normal and pathological samples. That is, $\hat{\Sigma} = (1/N) \sum_{l=1}^{N} (X_l - \widehat{M})(X_l - \widehat{M})^T$, where X_l, $l = 1, 2, \ldots, N$ are the feature vectors in the design set of cardinality N and $\widehat{M}$ is the gross sample mean feature vector. In the

Bayes sense, the linear classifier is optimum only for the normal distribution with equal covariance matrices [43]. Although, the assumption of equal covariance matrices might not be plausible in reality, the simplicity of the classifier compensates for any potential loss in accuracy other classifiers (e.g., SVMs) might deliver. Indeed, (4) requires only $\hat{\Sigma}$ and $\widehat{M}_i$, $i = 1, 2$ to be estimated from the design set. However, it should be stressed that no linear classifier performs well, when the distributions are not separated by the mean-difference, but are separated by the covariance-difference. In the latter case, one has to adopt a more complex classifier, for example, a quadratic one.

2.1. ROC Curve without Reject Option. The decisions taken by the linear classifier (4) for all test samples yield the following measures, which are functions of the threshold t:

(i) *true positive rate* (TP), also called *sensitivity* or *probability of detection* P_D, which is defined as the ratio between pathological samples correctly classified and the total number of pathological samples;

(ii) *false negative rate* (FN), also called *probability of miss*, which is defined as the ratio between pathological samples wrongly classified and the total number of pathological samples;

(iii) *true negative rate* (TN), also called *specificity*, which is defined as the ratio between normal samples correctly classified and the total number of normal samples;

(iv) *false positive rate* (FP) also known as *probability of false alarm* P_{FA}, which is defined as the ratio between normal samples wrongly classified and the total number of normal samples.

By varying the threshold, we obtain several operating points of the classifier, which can be represented through the *receiver operating characteristic* (ROC) curve, which is the plot of P_D (TP) versus P_{FA} (FP) having t as an implicit parameter. The ROC is always a concave upwards curve [50]. If a single figure of merit out of a ROC curve is sought, the most commonly used figure of merit is the area under the ROC curve. An ideal classifier would have a unit area under the ROC curve. Besides the visualization of classifier performance, the ROC curve can be used to select the most appropriate decision threshold for a particular application [47]. In this case, one has to resort to the costs c_{ij}, $i, j = 1, 2$, shown in the upper two rows in Table 1. Clearly, c_{12} and c_{21} are related to a false negative and a false positive classification, while c_{11} and c_{22} refer to the costs of true negative and true positive classifications. A particular operating point $(P_{FA}(t), P_D(t))$ at threshold t is associated to the expected cost [47]:

$$
\begin{aligned}
EC(t) = &P_1(c_{21} - c_{11})P_{FA}(t) + P_2(c_{22} - c_{12})P_D(t) \\
&+ P_1 c_{11} + P_2 c_{12}
\end{aligned}
\tag{5}
$$

which defines a set of straight lines with slope

$$
\alpha = -\frac{P_1}{P_2} \frac{c_{21} - c_{11}}{c_{22} - c_{12}}
\tag{6}
$$

TABLE 1: Costs for voice pathology detection with reject option.

Detector's decision	Actual diagnosis	
	Normal (1)	Pathological (2)
Normal (1)	c_{11}	c_{12}
Pathological (2)	c_{21}	c_{22}
Reject	c_{R1} (CRN)	c_{R2} (CRP)

on the $(P_{FA}(t), P_D(t))$ plane. Among these lines the one touches the ROC curve determines the best operating point, that is, the threshold that minimizes the expected cost. If the ROC curve has been obtained by means of a parametric model, it is a smooth curve and the best operating point is where the line is tangent to the ROC curve [50]. When the ROC curve is defined with respect to a finite number of experimental measurements connected with straight lines, the optimal operating point can be determined by the point where a line with slope α touches the ROC curve moving downwards from the top left corner of the (P_{FA}, P_D) plane [51]. Such point lies on the *ROC convex hull*. That is, the smallest convex set containing the points of the ROC curve [47].

3. Dichotomizers with Reject Option

Given X, the conditional error (or risk) for the Bayes classifier for minimum error (1) is

$$r(X) = \min\{P_1 p_1(X), P_2 p_2(X)\}. \qquad (7)$$

When $r(X)$ is close to 0.5, decision-making can be postponed by introducing a *reject test*. By setting a threshold θ for $r(X)$, the reject region is defined as [43]

$$r(X) \geq \theta \iff -\ln\frac{1-\theta}{\theta} + \ln\frac{P_1}{P_2}$$
$$\qquad (8)$$
$$\leq h(X) \leq \ln\frac{1-\theta}{\theta} + \ln\frac{P_1}{P_2}.$$

Thus whenever (8) is satisfied, the sample X is rejected. That is, no decision is taken by the classifier and further advice is requested by a medical doctor in the context of the application discussed in the paper. Samples in Ω_1 satisfying $h(X) > \ln((1-\theta)/\theta) + \ln(P_1/P_2)$ are misclassified (FP). Similarly, samples in Ω_2 satisfying $h(X) < -\ln((1-\theta)/\theta) + \ln(P_1/P_2)$ are misclassified (FN). Equation (8) suggests to modify the linear classifier decision rule (4) by introducing two thresholds t_1 and t_2 with $t_1 \leq t_2$ as follows:

$$X \in \Omega_1(N) \quad \text{if } \hat{h}(X) < t_1,$$
$$X \in \Omega_2(P) \quad \text{if } \hat{h}(X) > t_2, \qquad (9)$$
$$X \text{ is rejected} \quad \text{if } t_1 \leq \hat{h}(X) \leq t_2.$$

Obviously, (9) suggests that although the probability of rejection is a fraction of the test samples, the probability of false alarm and the probability of detection is now a fraction

of the test samples, which are not being rejected. That is, the denominators in the estimates of the just mentioned probabilities are now different than those without rejection.

In a sample-based approach, we may set $t_1 = t - \vartheta$ and $t_2 = t + \vartheta$, where t admits values uniformly spaced in the interval $[h_{\min}, h_{\max}]$ with $h_{\min} = \min_{X \in (\Omega_1 \cup \Omega_2)}\{\hat{h}(X)\}$ and $h_{\max} = \max_{X \in (\Omega_1 \cup \Omega_2)}\{\hat{h}(X)\}$, while $\vartheta = \gamma\Delta t$, where Δt is the step increment of t and γ is a small integer. However, such a choice does not harm the validity of the analysis following for generic (asymmetric) thresholds t_1 and t_2 [47]. Let $\mathcal{T}$ the set of discrete thresholds determined by the just described procedure for t. One may set $t_1 \in \mathcal{T}$ and $t_2 \in \mathcal{T}$ so that $t_2 > t_1$.

3.1. ROC Curve with Reject Option. When a reject option is introduced in the classifier design, the costs for rejection should be inserted in the last row of Table 1. The optimal values of t and ϑ (or γ) should be determined so that the following two conflicting requirements are fulfilled, namely classification error reduction and limited reject region in order to preserve as many correct classifications as possible. Following similar lines to [47], it can be shown that the expected cost associated with the classification (9) is now a function of two variables and is given by

$$\text{EC}(t, \vartheta) = \epsilon_2(t + \vartheta) - \epsilon_1(t - \vartheta) + P_2 c_{12} + P_1 c_{11}, \qquad (10)$$

where

$$\epsilon_1(t - \vartheta) = P_2(c_{12} - c_{R2})P_D(t - \vartheta) + P_1(c_{11} - c_{R1})P_{FA}(t - \vartheta),$$
$$\epsilon_2(t + \vartheta) = P_2(c_{22} - c_{R2})P_D(t - \vartheta) + P_1(c_{21} - c_{R1})P_{FA}(t - \vartheta). \qquad (11)$$

The optimal t and ϑ satisfy $\nabla_{t,\vartheta}\text{EC}(t, \vartheta) = \mathbf{0}$. This is equivalent to

$$P_2(c_{22} - c_{R2})\frac{\partial P_D(t_2)}{\partial t_2} + P_1(c_{21} - c_{R1})\frac{\partial P_{FA}(t_2)}{\partial t_2}$$
$$- P_2(c_{12} - c_{R2})\frac{\partial P_D(t_1)}{\partial t_1} - P_1(c_{11} - c_{R1})\frac{\partial P_{FA}(t_1)}{\partial t_1} = 0,$$
$$P_2(c_{22} - c_{R2})\frac{\partial P_D(t_2)}{\partial t_2} + P_1(c_{21} - c_{R1})\frac{\partial P_{FA}(t_2)}{\partial t_2}$$
$$+ P_2(c_{12} - c_{R2})\frac{\partial P_D(t_1)}{\partial t_1} + P_1(c_{11} - c_{R1})\frac{\partial P_{FA}(t_1)}{\partial t_1} = 0, \qquad (12)$$

where the following change of variables has been made $t_1 = t - \vartheta$ and $t_2 = t + \vartheta$. By adding and subtracting by parts the two equations in the set (12), we arrive at

$$P_2(c_{22} - c_{R2})\frac{\partial P_D(t_2)}{\partial t_2} + P_1(c_{21} - c_{R1})\frac{\partial P_{FA}(t_2)}{\partial t_2} = 0,$$
$$P_2(c_{12} - c_{R2})\frac{\partial P_D(t_1)}{\partial t_1} + P_1(c_{11} - c_{R1})\frac{\partial P_{FA}(t_1)}{\partial t_1} = 0. \qquad (13)$$
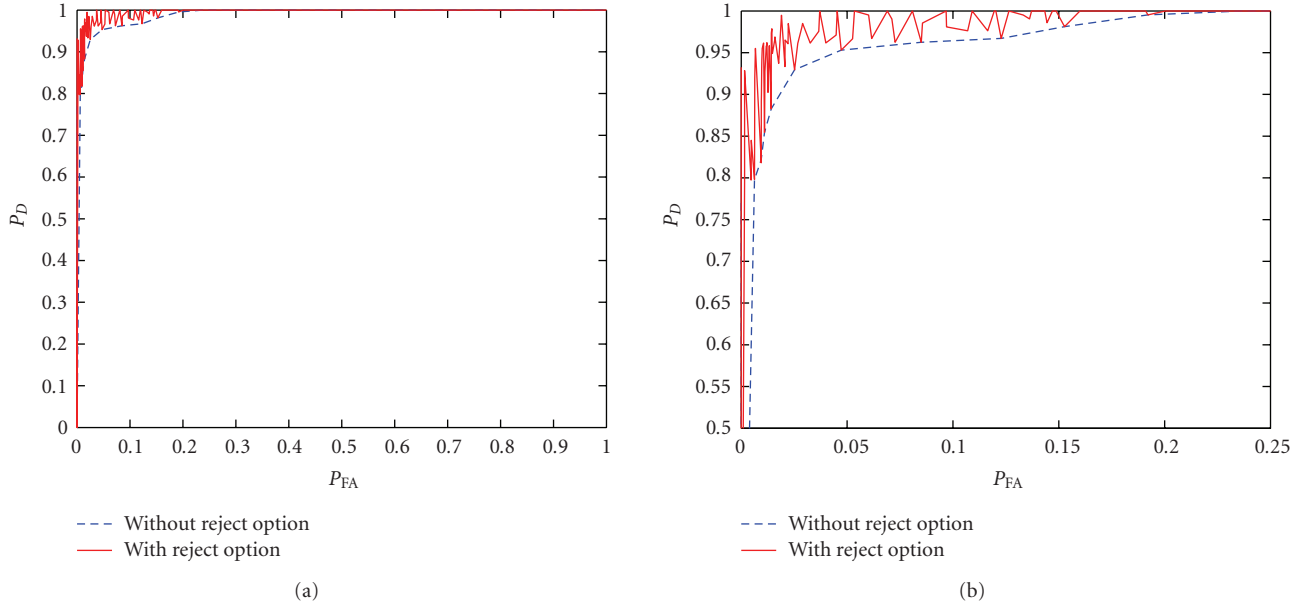
FIGURE 1: (a) Experimental ROC curves of the linear classifier tested for vocal fold paralysis detection in men without reject option (dashed line) and with reject option (solid line). (b) Zoom in the ROC curves.

The set of equations (13) defines two straight lines with slopes

$$\alpha_1 = -\frac{P_1}{P_2}\frac{c_{21} - c_{R1}}{c_{22} - c_{R2}}, \tag{14}$$

$$\alpha_2 = -\frac{P_1}{P_2}\frac{c_{11} - c_{R1}}{c_{12} - c_{R2}} \tag{15}$$

on the plane of P_{FA} and P_D. Equations (14) and (15) are valid for generic t_1 and t_2. The set of equations (13) suggests that the straight lines of slope α_1 and α_2 should touch the convex hull of the ROC curve without reject option at two distinct points having implicit parameters t_1 and t_2 such that $t_1 < t_2$. Each of these distinct points can be found by means of a simple search of the edges of the ROC convex hull derived without the reject option [47]. Having found t_1 and t_2, the set of equations $t_1 = t - \vartheta$ and $t_2 = t + \vartheta$ is then solved for t and ϑ. Clearly, the just derived estimates of t and ϑ are initial ones, because they depend on the convex hull resolution of the ROC curve without rejection estimated from the threshold values $t \in \mathcal{T}$. The initial estimates of t and ϑ can be corrected, when the operating point they define lies inside the convex hull of the ROC curve with rejection. Since the probability of false alarm and the probability of detection in the latter ROC curve are fractions of the test samples, which are not being rejected, the lines of slope α given by (6) should touch the convex hull of the ROC curve with rejection at the optimal operating point. The values of t and ϑ of the aforementioned optimal operating point are better estimates than the initial ones. If the initial estimates of t and ϑ define an operating point outside the convex hull of the ROC curve with rejection, then no further correction is needed, because such an operating point defines a new vertex of the convex hull linked by two new edges with the

nearest vertices already included in the available convex hull. Obviously, the new vertex will be the point where the lines of slope α touch the updated convex hull.

4. Datasets and Feature Extraction

The MEEI database was released in 1994 [37]. It contains over 1400 voice signals of approximately 700 subjects. Two different kinds of recordings were collected: the patients were called to articulate the sustained vowel "ah" (/a/) and to read the "rainbow passage" in each session. The database contains recordings of vowel "ah" (53 normal and 657 pathological utterances) and continuous speech (53 normal and 661 pathological utterances). The discussion is focused on the sustained vowel recordings and first results on "rainbow passage" recordings will be reported. The recordings were performed in matching acoustic conditions, using Kays Computerized Speech Lab. Each subject was asked to produce a sustained phonation of vowel "ah" at a comfortable pitch and loudness for at least 3 seconds. The process was repeated three times for each subject, and a speech pathologist chose the best sample for the database. The recordings of the sustained vowel were made at a sampling rate of 25 KHz for patients and 50 KHz for the healthy subjects. In the latter case, the sampling rate was reduced to 25 KHz by down-sampling. The normal voice recordings are about 5 seconds long, whereas the pathological ones are about 3 seconds long. The major asset of the MEEI database is the clinical assessment of the subjects as well as the availability of subjects' personal details. However, there are several drawbacks that are carefully identified in [21].

Due to the inherent differences in the speech production system of male and female subjects, it makes sense to deal
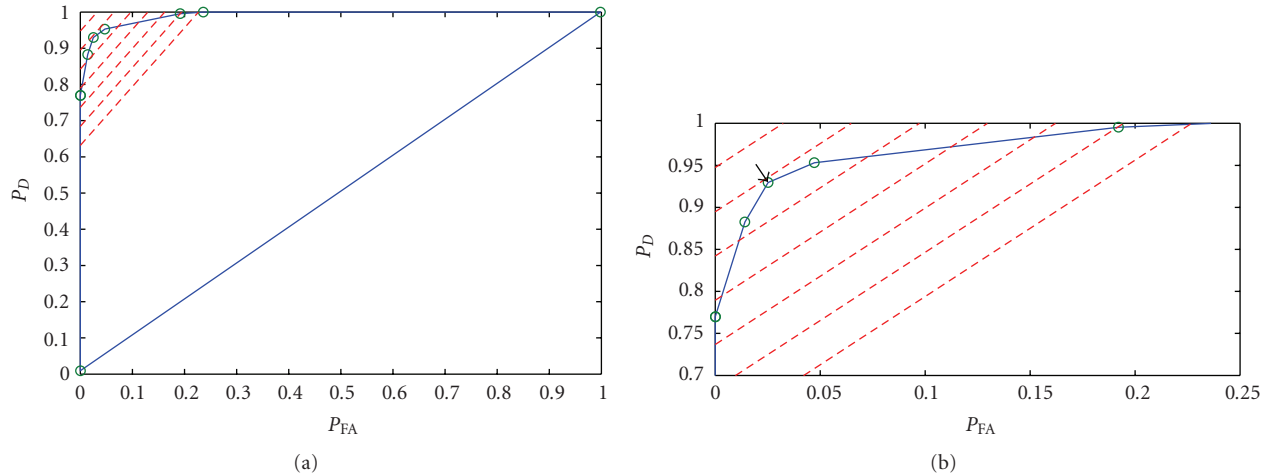
FIGURE 2: (a) Convex hull of the experimental ROC curve of the linear classifier without reject option (solid line) with the level lines of slope α (dashed lines) overlaid. (b) Zoom in (a): the arrow points to the optimal operating point $(P_{FA}, P_D) = (0.0252, 0.9296)$.

with disordered speech detection separately for each gender. Two experiments are conducted. The first experiment concerns vocal fold paralysis detection and the dataset comprises recordings from 21 males aged 26 to 60 years, who were medically diagnosed as normal, and another 21 males aged 20 to 75 years, who were medically diagnosed with vocal fold paralysis. The second experiment concerns vocal fold edema detection, where 21 females aged 22 to 52 years, who were medically diagnosed as normal, and another 21 females aged 18 to 57 years, who were medically diagnosed with vocal fold edema served as subjects. The subjects might suffer from other diseases too, such as hyperfunction, ventricular compression, atrophy, teflon granuloma, and so forth. Although a multi-label classification framework would be more appropriate, we will assume a sort of tying in this paper by ignoring the other connotations, so that enough design and test samples are available for our study. Multi-label classification is left for future research. However, the linear classifier studied in the paper requires only the estimation of the class-conditional mean vectors and the gross dispersion matrix. Accordingly, the number of adjustable parameters is not high.

As in [29, 32], 14 LPCs are extracted for each speech frame. The speech frames have a duration of 20 ms and neighboring frames do not overlap. The rectangular window is used to extract the speech frames. By varying the number of LPCs from 14 to 30, we have found that the probability of correct classification for both voice pathologies does not improve so much to justify linear prediction analysis of higher order than the 14th. On the contrary, more LPCs than 14 are found to frequently deteriorate the probability of correct classification.

In the first experiment, the sample set consists of 4236 14-dimensional feature vectors (i.e., samples) of which 3171 samples were extracted from normal speech utterances of the sustained vowel "ah" and the remaining 1065 samples were extracted from pathological speech uttered by male speakers. In the second experiment, the sample set consists of 4199

TABLE 2: Arithmetic values of the costs employed for voice pathology detection with reject option.

Detector's decision	Actual diagnosis	
	Normal (1)	Pathological (2)
Normal (1)	−1	10
Pathological (2)	5	−1
Reject	1	2

14-dimensional feature vectors of which 3096 samples were extracted from normal speech utterances of the sustained vowel "ah" and the remaining 1103 samples were extracted from pathological speech uttered by female speakers. For each experiment, first experimental results using utterances of "rainbow passage" are also reported.

5. Experimental Results

The assessment of the linear classifier for detecting vocal fold paralysis in men and vocal fold edema in women either with or without reject option is based on the ROC curve. 80% of the samples have been used in classifier design, and the remaining 20% of the samples has been used for testing the classifier. The classifier design aims at estimating the parameters appearing in (4). The costs depicted in Table 2 have been used in the study of ROC curves. The negative sign for true positives and true negatives should be interpreted as a gain. The assignment of a higher cost for false negatives (misses) than false positives (false alarms) is easily understood. The costs c_{R2} (CRP) and c_{R1} (CRN) are chosen so that the inequality

$$\frac{c_{11} - c_{R1}}{c_{12} - c_{R2}} > \frac{c_{21} - c_{R1}}{c_{22} - c_{R2}} \tag{16}$$
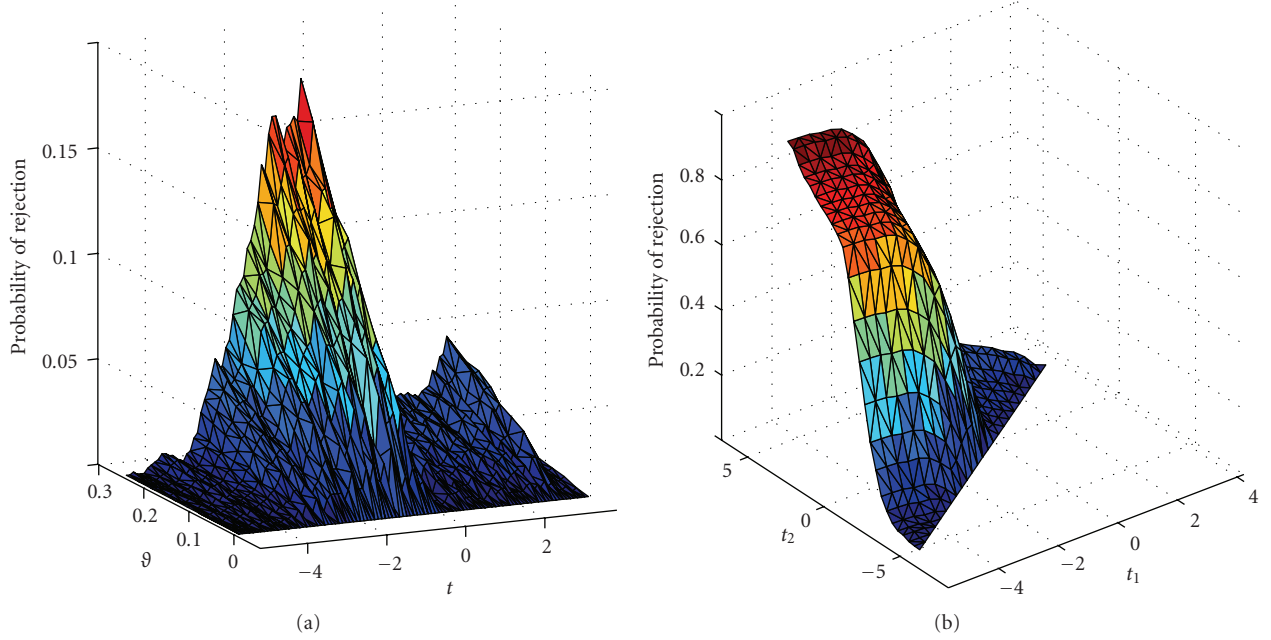
holds [47]. A design strategy is as follows.

(a)

(b)

FIGURE 3: Probability of rejection in vocal fold paralysis detection as a function of (a) t and ϑ, (b) $t_1, t_2 \in \mathcal{T}$ with $t_2 \geq t_1$.
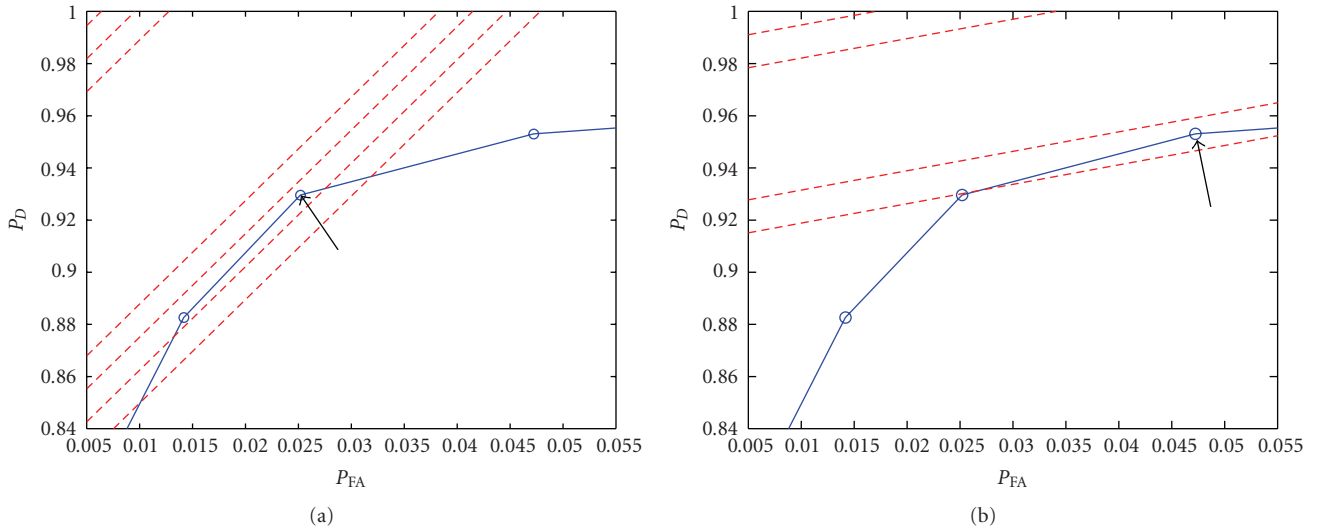

(a)

(b)

FIGURE 4: (a) Zoom in the convex hull of the ROC without reject option (solid line); the level lines of slope α_1 (dashed lines) are overlaid. The arrow points to the optimal operating point $(P_{\mathrm{FA}}(t_2), P_D(t_2)) = (0.0252, 0.9296)$. (b) Zoom in the convex hull of the ROC without reject option (solid line); the level lines of slope α_2 (dashed lines) are overlaid. The arrow points to the optimal operating point $(P_{\mathrm{FA}}(t_1), P_D(t_1)) = (0.0472, 0.9531)$.

(1) Choose $c_{22} < c_{R2} < c_{12}$, for example, $c_{R2} = 2$.

(2) Let $\eta = (c_{12} - c_{R2})/(c_{R2} - c_{22}) > 0$, for example, $\eta = 1$.

(3) Then, $c_{R1} < (c_{21}\eta + c_{11})/\eta + 1$, for example, $c_{R1} < 4.5$.

In addition, c_{R1} should be chosen so that the straight lines of slope α_1 and α_2 touch the convex hull of the ROC curve without reject option at two distinct points in order the reject option to be meaningful. The choice $c_{R1} = 1$ satisfies both requirements. However, any other assignment stemming from the just described strategy could also be used.

5.1. Vocal Fold Paralysis in Men. The experimental ROC curves of the linear classifier without reject option (4) and with reject option (9), that were derived by counting classifier decisions, are shown in Figure 1.

In order to obtain a better insight into the detection, first the convex hull of the ROC curve without the reject option is plotted in Figure 2(a). In the same figure, several parallel level lines $P_D(t) = \alpha P_{\mathrm{FA}}(t) + \beta(t)$ are overlaid. Clearly, one of these lines passes through the ideal operating point $(P_{\mathrm{FA}}(t), P_D(t)) = (0, 1)$. The intercept of this line
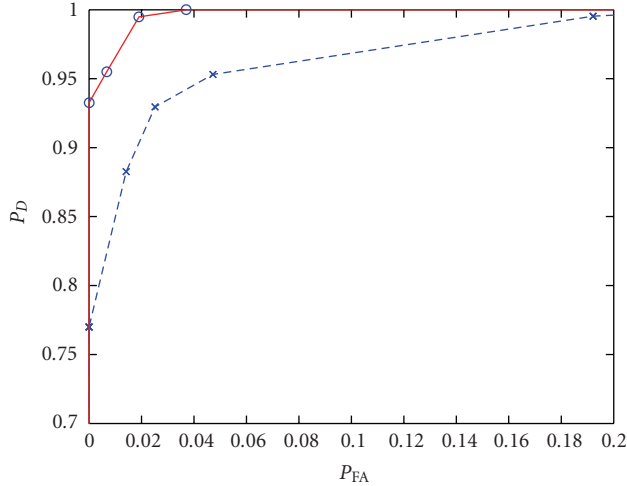
FIGURE 5: Zoom in the ROC convex hulls with reject option (solid line) and without reject option (dashed line).
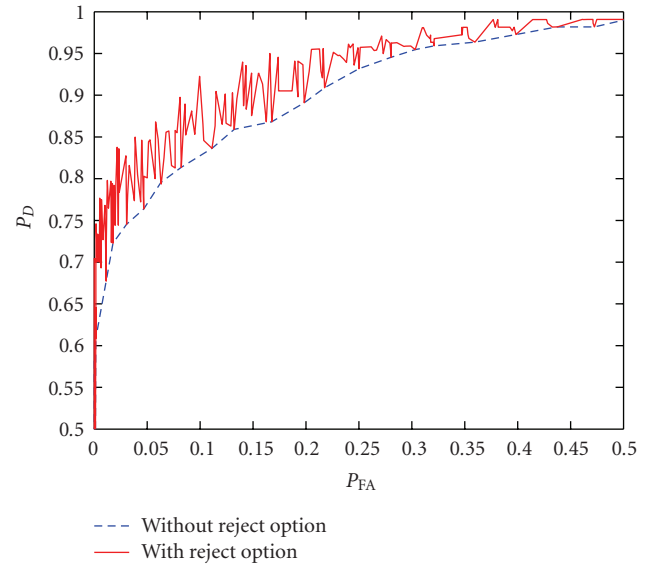

- - - Without reject option
— With reject option

FIGURE 6: Zoom in the experimental ROC curves of the linear classifier applied to vocal fold edema detection in women without reject option (dashed line) and with reject option (solid line).

is $\beta(t)|_{\{t:P_{FA}(t)=0, P_D(t)=1\}} = 1$. Accordingly, to produce the set of parallel lines one has to uniformly vary $\beta \in [0, 1]$. The inspection of Figure 2(b) reveals the optimal operating point $(P_{FA}(t), P_D(t)) = (0.0252, 0.9296)$, where the level lines touch the ROC convex hull. Indeed, the line above that touching the ROC curve does not determine any feasible point for the classifier, although it exhibits a lower expected cost, while the line below intersects the ROC curve in at least two points, but at a greater expected cost. The easiest method to identify the optimal point is the visual inspection of the graph. However, since the vertices of the convex hull have already been determined, one has to insert the associated $(P_{FA}(t), P_D(t))$ into (5), sort the vertices in increasing order of the expected cost, and read the operating point that yields the minimum expected cost. Alternatively, one may search the edges of the ROC convex hull as is suggested in [47]. All these methods have been successfully tested in all experiments conducted.

The introduction of the reject option in (9) induces the probability of rejection, which is plotted in Figure 3 as a function of t_1 and t_2 when the costs shown in Table 2 are used. Figure 3(a) depicts the probability of rejection as a function of t and ϑ. In particular, $t \in \mathcal{T}$ and 10 equally spaced values of $\vartheta \in [0, 3\Delta t]$ were defined. As expected, the largest probability of rejection (i.e., 0.1804) occurs for $t = -0.7330$ and $\vartheta = 0.2434$ yielding thresholds t_1 and t_2 in the middle of their domain $\mathcal{T}$. The probability of rejection for $t_1, t_2 \in \mathcal{T}$ with $t_2 \geq t_1$ is plotted in Figure 3(b). It is seen that the generic rejection region may yield large probabilities of rejection leaving very few test samples to be processed by the classifier. On the contrary, much fewer test samples should be submitted to a clinician for further screening, if t_1, t_2 are set equal to $t \pm \vartheta$.

In Figure 4(a), the convex hull of the ROC without rejection is plotted along with the level lines having slope α_1 given by (14). The points that define the ROC convex hull are indicated by markers. The level lines touch the ROC convex hull at the operating point $(P_{FA}(t_2), P_D(t_2)) =$

$(0.0252, 0.9296)$. The level lines having slope α_2 given by (15) touch the convex hull of the ROC without rejection at the operating point $(P_{FA}(t_1), P_D(t_1)) = (0.0472, 0.953)$, as can be seen in Figure 4(b). The implicit thresholds associated with the two operating points are $t_1 = -0.2822$ and $t_2 = -0.1920$. Indeed, the reject option is useful in the middle of the domain of thresholds $\mathcal{T}$. By applying the procedure described in Section 3.1, the associated probabilities of false alarm and detection with reject option at the optimal operating point are found to be 0.01904 and 0.99484. It is seen that the introduction of rejection has improved the probability of detection by 6.59% for probability of false alarm fixed to approximately 2%. The classification accuracy with reject option at the operating point under discussion is measured 98.47%, that is 2.13% higher than that measured without rejection. The confidence interval for the classification accuracy can be estimated as in [21], that is,

$$CI = \pm z_{1-\delta/2}\sqrt{\frac{q(1-q)}{N}}, \qquad (17)$$

where $z_{1-\delta/2}$ is the standard Gaussian percentile for confidence level $100(1-\delta)\%$ (e.g., for $\delta = 0.05$, $z_{1-\delta/2} = z_{0.975} = 1.967$), q is the experimentally measured classification accuracy, and N is the number of samples. In our case, for $N = 847$ and $q = 0.96863$, (17) yields 0.83%, which indicates that the just mentioned improvement is statistically significant at 95% level of significance. If c_{R1} is set equal to -1 (i.e., a gain is introduced for rejecting normal subjects), which is a permissible policy according to the cost assignment methodology described previously, and all other costs are left intact, the probability of correct classification at the best operating point increases to 98.59%, which yields a statistically significant improvement at the same level of significance (CI = 0.7954%). At the latter operating point,
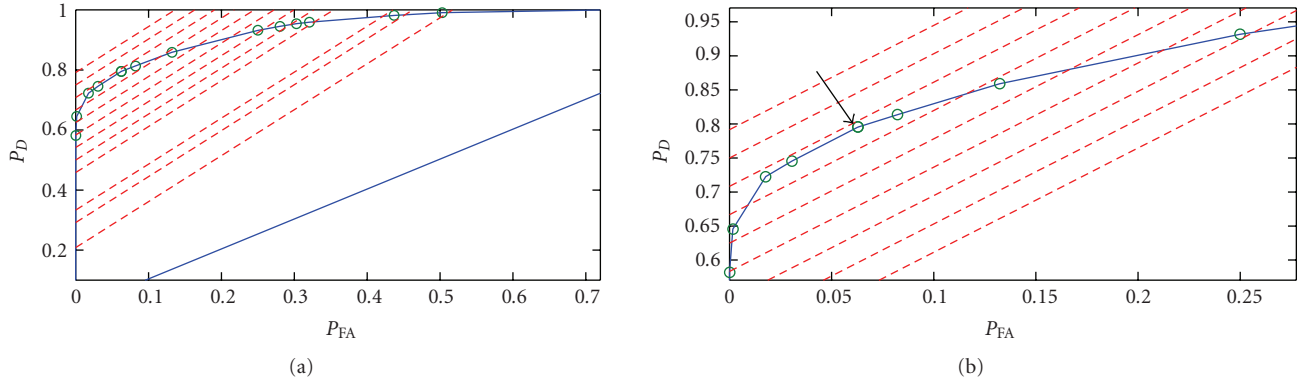
(a)

(b)

FIGURE 7: (a) Convex hull of the experimental ROC curve of the linear classifier without reject option (solid line) with the level lines of slope α (dashed lines) overlaid. (b) Zoom in (a), the arrow points to the optimal operating point $(P_{\mathrm{FA}}, P_D) = (0.0629, 0.7955)$.

we have $P_{\mathrm{FA}} = 0.0172$ and $P_D = 0.994709$, when the reject option is enabled.

The superiority of the linear classifier with reject option is demonstrated in Figure 5, where the convex hull of the ROC curves with reject option (solid line) and without reject option (dashed line) are plotted only. It is self-evident that the area of the convex hull for the ROC with reject option is greater than that without reject option. The area of the convex hull is correlated with the area under the ROC that is frequently used as an objective figure of merit. In particular, the area under the ROC was measured to 0.9868 without rejection and 0.9951 with rejection option, when $t_1 = t - \vartheta$ and $t_2 = t + \vartheta$.

The same procedure has been applied to a set of 5049 test feature vectors extracted from utterances of "rainbow passage." At the optimal operating point with respect to the costs of Table 2 the classifier without reject option yields $P_{\mathrm{FA}} = 0.477227$ and $P_D = 0.9358$ and its accuracy is 72.93%. The introduction of the reject option yields at the optimal operating point $P_{\mathrm{FA}} = 0.0686$ and $P_D = 0.91875$, while the probability of correct classification increases to 92.45%. It is seen that the reject option reduces drastically the probability of false alarm by approximately 40% at the same probability of detection. Needless to say that the improvement in classification accuracy is statistically significant.

5.2. Vocal Fold Edema in Women.

The experimental ROC curves of the linear classifier without reject option (4) and with reject option (9) with the cost assignment shown in Table 2 were derived by counting classifier decisions are plotted in Figure 6.

The convex hull of the ROC curve without reject option is plotted in Figure 7. In the same figure, a set of parallel level lines having slope given by (6) is overlaid and the points that define the ROC convex hull are indicated by markers. If the costs shown in Table 2 are employed, the minimum expected cost is found for the threshold that yields the operating point $(P_{\mathrm{FA}}(t), P_D(t)) = (0.0629, 0.7955)$, where the level lines touch the ROC convex hull.

The introduction of the reject option in (9) induces the probability of rejection, which is plotted in Figure 8 as a
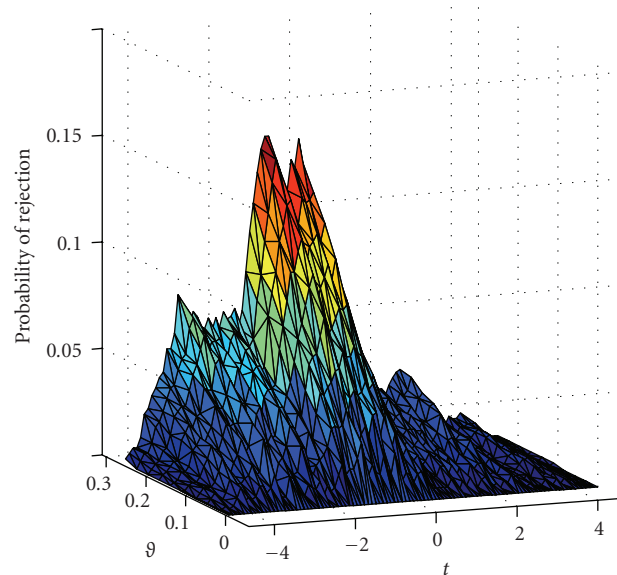

FIGURE 8: Probability of rejection as a function of (t_1, t_2) for vocal fold edema detection.

function of t and ϑ. 100 equally spaced values in the range $[h_{\min}, h_{\max}]$ were taken for t and 10 equally spaced values of $\vartheta \in [0, 3\Delta t]$ were defined as previously in vocal fold paralysis. As expected, the larger probability of rejection occurs in the middle of the domain of $t \pm \vartheta$.

In Figure 9(a), the convex hull of the ROC without rejection is plotted along with the level lines having slope α_1 given by (14). The points that define the ROC convex hull are indicated by markers. The level lines touch the ROC convex hull at the operating point $(P_{\mathrm{FA}}(t_2), P_D(t_2)) = (0.0177, 0.7227)$. The level lines of slope α_2 given by (15) touch the convex hull of the ROC without rejection at the operating point $(P_{\mathrm{FA}}(t_1), P_D(t_1)) = (0.1322, 0.8590)$, as is demonstrated in Figure 9(b). These operating points correspond to $t_1 = -0.2643$ and $t_2 = 0.2937$. By applying the procedure described in Section 3.1, the associated probabilities of false alarm and detection with reject option are found
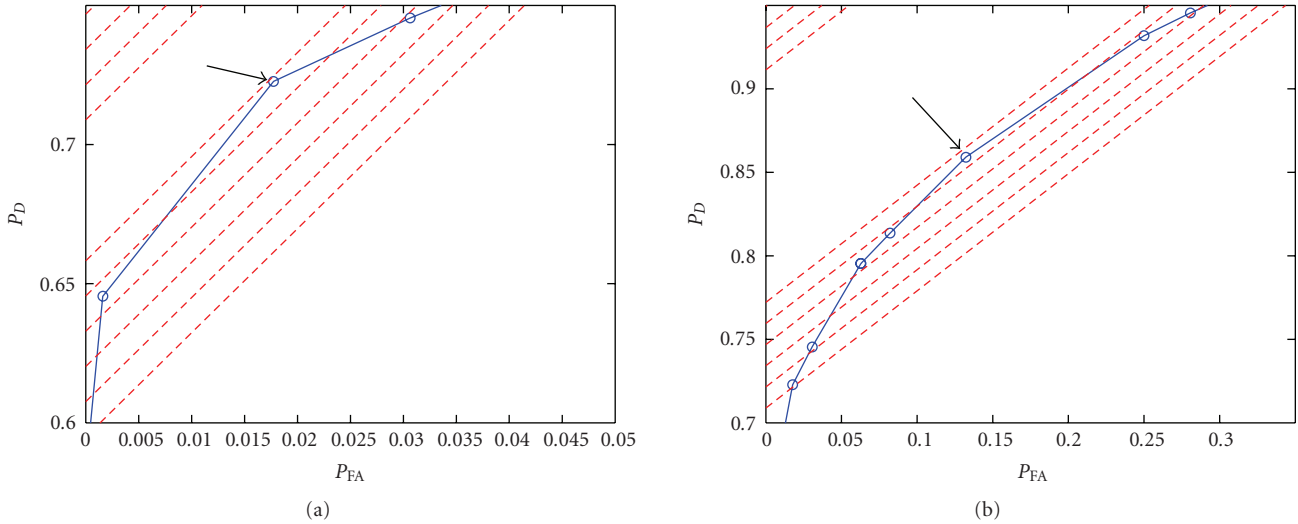
FIGURE 9: (a) Zoom in the convex hull of the ROC without reject option (solid line); The level lines of slope α_1 (dashed lines) are overlaid. The arrow points to the optimal operating point $(P_{\mathrm{FA}}(t_2), P_D(t_2)) = (0.0177, 0.7227)$. (b) Zoom in the convex hull of the ROC without reject option (solid line); the level lines of slope α_2 (dashed lines) are overlaid. The arrow points to the optimal operating point $(P_{\mathrm{FA}}(t_1), P_D(t_1)) = (0.1322, 0.8590)$.
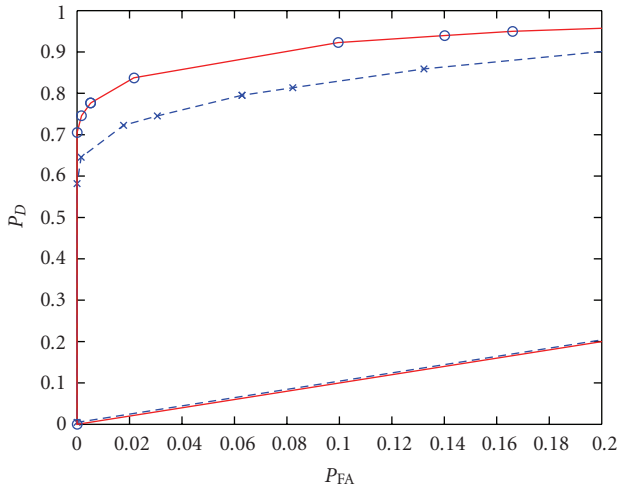

FIGURE 10: Zoom in the ROC convex hulls with reject option (solid line) and without reject option (dashed line).

to be 0.02003 and 0.836842, respectively. The classification accuracy with reject option at the best operating point, when the costs of Table 2 are used, is measured 94.316%. That is, 4.316% higher than that measured without rejection. The confidence interval for the classification accuracy predicted by (17) for $N = 840$ and $q = 0.94316$ is 1.57%, which indicates that the just mentioned improvement of 4.316% is statistically significant at 95% level of significance. By fixing the probability of detection to 83.64%, the reject option is found to reduce the probability of false alarm by 9.12%.

The superiority of the linear classifier with reject option is demonstrated in Figure 10, where the convex hull of the ROC curves with reject option (solid line) and without reject

option (dashed line) are plotted only. It is self-evident that the area of the convex hull for the ROC with reject option is greater than that without reject option. In particular, the area under the ROC increases from 0.9458 to 0.96 with the introduction of the reject option.

The same procedure has been applied to a set of 3365 test feature vectors extracted from utterances of "rainbow passage." At the optimal operating point with respect to the costs of Table 2 the classifier without reject option yields $P_{\mathrm{FA}} = 0.5965$ and $P_D = 0.8959$ and its probability of correct classification is 64.96%. The introduction of the reject option yields at the optimal operating point $P_{\mathrm{FA}} = 0.5228$ and $P_D = 0.8853$, while the accuracy increases to 68.8%. It is seen that the reject option reduces the probability of false alarm by approximately 7.3% at the same probability of detection. The improvement of 3.9% in classification accuracy is statistically significant at 95% level of significance (CI = 1.57%).

6. Conclusions

The reject option has been shown to improve the accuracy of a linear classifier in detecting vocal fold paralysis for male patients as well as detecting vocal fold edema for female ones than that obtained without reject option. Moreover, the reported improvements are shown to be statistically significant at 95% confidence level. In addition, the linear classifier with reject option outperforms the previously employed classifiers in [29, 32] to detect the aforementioned voice pathologies under exactly the same experimental protocol. Future research will address the introduction of reject option in the design of the Bayes classifier, when Gaussian mixture models approximate the class conditional probability density functions of the linear prediction coefficients extracted from continuous speech.

References

[1] C. Manfredi, "Voice models and analysis for biomedical applications," *Biomedical Signal Processing and Control*, vol. 1, no. 2, pp. 99–101, 2006.

[2] F. Quek, M. Harper, Y. Haciahmetoglou, L. Chen, and L. O. Ramig, "Speech pauses and gestural holds in parkinson's disease," in *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP '02)*, pp. 2485–2488, Denver, Colo, USA, September 2002.

[3] L. Will, L. O. Ramig, and J. L. Spielman, "Application of lee silverman voice treatment (LSVT) to individuals with multiple sclerosis, ataxic dysarthria, and stroke," in *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP '02)*, pp. 2497–2500, Denver, Colo, USA, September 2002.

[4] P. Enderby and L. Emerson, *Does Speech and Language Therapy Work?* Singular Publications, 1995.

[5] R. P. Schumeyer and K. E. Barner, "Effect of visual information on word initial consonant perception of dysarthric speech," in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, vol. 1, pp. 46–49, Philadelphia, Pa, USA, October 1996.

[6] K. Mády, R. Sader, A. Zimmermann, et al., "Assessment of consonant articulation in glossectomee speech by dynamic MRI," in *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP '02)*, pp. 961–964, Denver, Colo, USA, September 2002.

[7] R. Schwarz, U. Hoppe, M. Schuster, T. Wurzbacher, U. Eysholdt, and J. Lohscheller, "Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1099–1108, 2006.

[8] V. Parsa and D. G. Jamieson, "Interactions between speech coders and disordered speech," *Speech Communication*, vol. 40, no. 7, pp. 365–385, 2003.

[9] M. S. Hawley, P. Green, P. Enderby, S. Cunningham, and R. K. Moore, "Speech technology for e-inclusion of people with physical disabilities and disordered speech," in *Proceedings of the 9th European Conference on Speech Communication and Technology (INTERSPEECH '05)*, pp. 445–448, Lisbon, Portugal, September 2005.

[10] F. Plante, H. Kessler, B. Cheetham, and J. Earis, "Speech monitoring of infective laryngitis," in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, vol. 2, pp. 749–752, Philadelphia, Pa, USA, October 1996.

[11] E. J. Wallen and J. H. L. Hansen, "Screening test for speech pathology assessment using objective quality measures," in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, vol. 2, pp. 776–779, Philadelphia, Pa, USA, October 1996.

[12] M. N. Vieira, F. R. McInnes, and M. A. Jack, "Robust F0 and jitter estimation in pathological voices," in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, vol. 2, pp. 745–748, Philadelphia, Pa, USA, October 1996.

[13] P. Mitev and S. Hadjitodorov, "Fundamental frequency estimation of voice of patients with laryngeal disorders," *Information Sciences*, vol. 156, no. 1-2, pp. 3–19, 2003.

[14] H. Weiping, W. Xiuxin, and P. Gómez, "Robust pitch extraction in pathological voice based on wavelet and cepstrum," in *Proceedings of the 12th European Signal Processing Conference (EUSIPCO '04)*, pp. 297–300, Vienna, Austria, September 2004.

[15] L. Deng, X. Shen, D. Jamieson, and J. Till, "Simulation of disordered speech using a frequency-domain vocal tract model," in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, vol. 2, pp. 768–771, Philadelphia, Pa, USA, October 1996.

[16] B. Gabelman and A. Alwan, "Analysis by synthesis of FM modulation and aspiration noise components in pathological voices," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02)*, vol. 1, pp. 449–452, Orlando, Fla, USA, May 2002.

[17] J. Hanquinet, F. Grenez, and J. Schoentgen, "Synthesis of disordered speech," in *Proceedings of the 9th European Conference on Speech Communication and Technology (INTERSPEECH '05)*, pp. 1077–1080, Lisbon, Portugal, September 2005.

[18] V. Parsa and D. G. Jamieson, "Acoustic discrimination of pathological voice: sustained vowels versus continuous speech," *Journal of Speech, Language, and Hearing Research*, vol. 44, no. 2, pp. 327–339, 2001.

[19] A. McAllister, "Acoustic, perceptual and physiological studies of ten-year-old children's voices," *Speech, Music and Hearing Quarterly Progress and Status Report*, vol. 38, no. 1, 1997.

[20] V. Uloza, V. Saferis, and I. Uloziene, "Perceptual and acoustic assessment of voice pathology and the efficacy of endolaryngeal phonomicrosurgery," *Journal of Voice*, vol. 19, no. 1, pp. 138–145, 2005.

[21] N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomedical Signal Processing and Control*, vol. 1, no. 2, pp. 120–128, 2006.

[22] J. B. Alonso, J. de Leon, I. Alonso, and M. A. Ferrer, "Automatic detection of pathologies in the voice by HOS based parameters," *EURASIP Journal on Applied Signal Processing*, vol. 2001, no. 4, pp. 275–284, 2001.

[23] R. B. Reilly, R. Moran, and P. Lacy, "Voice pathology assessment based on a dialogue system and speech analysis," in *Proceedings of the of the AAAI Fall Symposium on Dialogue Systems for Health Communication*, pp. 104–109, Washington, DC, USA, October 2004.

[24] K. Shama, A. Krishna, and N. U. Cholayya, "Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 85286, 9 pages, 2007.

[25] P. Gómez, J. I. Godino, F. Rodríguez, et al., "Evidence of vocal cord pathology from the mucosal wave cepstral contents," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '04)*, vol. 5, pp. 437–440, Montreal, Canada, May 2004.

[26] J. B. Alonso, F. D. de Maria, C. M. Trevieso, and M. A. Ferrer, "Using nonlinear features for voice disorder detection," in *Proceedings of the 3rd International Conference on Non-Linear Speech Processing (NOLISP '05)*, pp. 94–106, Barcelona, Spain, 2005.

[27] M. Little, P. McSharry, I. Moroz, and S. Roberts, "Nonlinear, biophysically-informed speech pathology detection," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, vol. 2, pp. 1080–1083, Toulouse, France, May 2006.

[28] P. Kukharchik, I. Kheidorov, E. Bovbel, and D. Ladeev, "Speech signal processing based on wavelets and SVM for vocal tract pathology detection," in *Proceedings of the 3rd International*

Conference on Image and Signal Processing (ICISP '08), vol. 5099 of *Lecture Notes in Computer Science*, pp. 192–199, Springer, Cherbourg-Octeville, France, July 2008.

[29] M. Marinaki, C. Kotropoulos, I. Pitas, and N. Maglaveras, "Automatic detection of vocal fold paralysis and edema," in *Proceedings of the International Conference on Spoken Language Processing (ICSLP '04)*, pp. 537–540, Jeju, South Korea, October 2004.

[30] P. Gómez, F. Díaz, A. Álvarez, et al., "Principal component analysis of spectral perturbation parameters for voice pathology detection," in *Proceedings of the18th IEEE Symposium on Computer-Based Medical Systems (CBMS '05)*, pp. 41–46, Dublin, Ireland, June 2005.

[31] C. Peng, W. Chen, and B. Wan, "A preliminary study of pathological voice classification," in *Proceedings of the 7th IEEE International Conference on Computer and Information Technology (CIT '07)*, pp. 1106–1110, October 2007.

[32] E. Ziogas and C. Kotropoulos, "Detection of vocal fold paralysis and edema using linear discriminant classifiers," in *Proceedings of the 4th Helenic Conference on Advances in Artificial Intelligence (SETN '06)*, vol. 3955 of *Lecture Notes in Computer Science*, pp. 454–464, Springer, Heraklion, Greece, May 2006.

[33] B. G. A. Aguiar Neto, J. M. Fechine, S. C. Costa, and M. Muppa, "Feature estimation for vocal fold edema detection using short-term cepstral analysis," in *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering (BIBE '07)*, pp. 1158–1162, October 2007.

[34] C. Fredouille, G. Pouchoulin, J.-F. Bonastre, M. Azzarello, A. Giovanni, and A. Ghio, "Application of automatic speaker recognition techniques to pathological voice assessment (dysphonia)," in *Proceedings of the 9th European Conference on Speech Communication and Technology (EUROSPEECH '05)*, pp. 149–152, Lisbon, Portugal, September 2005.

[35] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 19–41, 2000.

[36] http://emedicine.medscape.com/article/863779-overview.

[37] *Massachusetts Eye and Ear Infirmary, Voice Disorders Database, Version 1.03*, Kay Elemetrics Corp., Lincoln Park, NJ, USA, 1994, CD-ROM.

[38] A. A. Dibazar, S. Narayanan, and T. W. Berger, "Feature analysis for automatic detection of pathological speech," in *Proceedings of the 25th IEEE Annual International Conference of the Engineering in Medicine and Biology*, vol. 1, pp. 182–183, 2002.

[39] V. Parsa, D. G. Jamieson, K. Stenning, and H. A. Leeper, "On the estimation of signal-to-noise ratio in continuous speech for abnormal voices," in *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP '02)*, pp. 2505–2508, Denver, Colo, USA, September 2002.

[40] J. Nayak and P. S. Bhat, "Identification of voice disorders using speech samples," in *Proceedings of the 10th IEEE International Conference on Convergent Technologies for Asia-Pasific Region (TENCON '03)*, vol. 3, pp. 951–953, 2003.

[41] R. A. Prosek, A. A. Montgomery, B. E. Walden, and D. B. Hawkins, "An evaluation of residue features as correlates of voice disorders," *Journal of Communication Disorders*, vol. 20, pp. 105–107, 1987.

[42] M. De Oliveira Rosa, J. C. Pereira, and M. Grellet, "Adaptive estimation of residue signal for voice pathology diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 1, pp. 96–104, 2000.

[43] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, Calif, USA, 2nd edition, 1990.

[44] X. Tang and W. Wang, "Dual-space linear discriminant analysis for face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. 1064–1068, 2004.

[45] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete Time Processing of Speech Signals*, MacMillan Publishing Company, New York, NY, USA, 1993.

[46] T. C. W. Landgrebe, D. M. J. Tax, P. Paclík, and R. P. W. Duin, "The interaction between classification and reject performance for distance-based reject-option classifiers," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 908–917, 2006.

[47] F. Tortorella, "A ROC-based reject rule for dichotomizers," *Pattern Recognition Letters*, vol. 26, no. 2, pp. 167–180, 2005.

[48] C. M. Santos-Pereira and A. M. Pires, "On optimal reject rules and ROC curves," *Pattern Recognition Letters*, vol. 26, no. 7, pp. 943–952, 2005.

[49] C. Marrocco, M. Molinara, and F. Tortorella, "An empirical comparison of ideal and empirical ROC-based reject rules," in *Proceedings of the 5th International Conference on Machine Learning and Data Mining (MLDM '07)*, vol. 4571 of *Lecture Notes in Computer Science*, pp. 47–60, 2007.

[50] H. L. V. Trees, *Detection, Estimation and Modulation Theory, Part I*, John Wiley & Sons, New York, NY, USA, 1968.

[51] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clinical Chemistry*, vol. 39, no. 4, pp. 561–577, 1993.