



Published in final edited form as:

*IEEE Trans Med Imaging*. 2008 November ; 27(11): 1556–1567. doi:10.1109/TMI.2008.928921.

## Application of Three-Class ROC Analysis to Task-Based Image Quality Assessment of Simultaneous Dual-Isotope Myocardial Perfusion SPECT (MPS)

**Xin He\*** [Member, IEEE],

Department of Radiology, Johns Hopkins School of Medicine, 601 N. Caroline Street, Baltimore, MD 21287 USA (e-mail: xinhe@jhmi.edu)

**Xiyun Song**, and

Philips Medical Systems, San Jose, CA 95134 USA

**Eric C. Frey** [Senior Member, IEEE]

Department of Radiology, Johns Hopkins School of Medicine, Baltimore, MD 21287 USA (e-mail: efrey@jhmi.edu)

### Abstract

The diagnosis of cardiac disease using dual-isotope myocardial perfusion SPECT (MPS) is based on the defect status in both stress and rest images, and can be modeled as a three-class task of classifying patients as having no, reversible, or fixed perfusion defects. Simultaneous acquisition protocols for dual-isotope MPS imaging have gained much interest due to their advantages including perfect registration of the  $^{201}\text{Tl}$  and  $^{99\text{m}}\text{Tc}$  images in space and time, increased patient comfort, and higher clinical throughput. As a result of simultaneous acquisition, however, crosstalk contamination, where photons emitted by one isotope contribute to the image of the other isotope, degrades image quality. Minimizing the crosstalk is important in obtaining the best possible image quality. One way to minimize the crosstalk is to optimize the injected activity of the two isotopes by considering the three-class nature of the diagnostic problem. To effectively do so, we have previously developed a three-class receiver operating characteristic (ROC) analysis methodology that extends and unifies the decision theoretic, linear discriminant analysis, and psychophysical foundations of binary ROC analysis in a three-class paradigm. In this work, we applied the proposed three-class ROC methodology to the assessment of the image quality of simultaneous dual-isotope MPS imaging techniques and the determination of the optimal injected activity combination. In addition to this application, the rapid development of diagnostic imaging techniques has produced an increasing number of clinical diagnostic tasks that involve not only disease detection, but also disease characterization and are thus multiclass tasks. This paper provides a practical example of the application of the proposed three-class ROC analysis methodology to medical problems.

### Index Terms

Dual-isotope myocardial perfusion SPECT (MPS); simultaneous acquisition; task-based image quality assessment; three-class Hotelling observer; three-class receiver operating characteristic (ROC) analysis

---

\*X. He is with the Department of Radiology, Johns Hopkins School of Medicine, 601 N. Caroline Street, Baltimore, MD 21287 USA (e-mail: xinhe@jhmi.edu).

Dual-Isotope myocardial perfusion SPECT (MPS) imaging using rest  $^{201}\text{Tl}$  chloride and stress  $^{99\text{m}}\text{Tc}$  sestamibi images has become one of the most commonly used clinical protocols for MPS imaging. The diagnosis of cardiac disease using dual-isotope MPS is based on the defect status in both the stress and rest images. It can be modeled as a three-class task of classifying patients as having no, fixed, or reversible defects. Simultaneous acquisition protocols for dual-isotope MPS imaging have gained much interest due to their advantages including perfect registration of the  $^{201}\text{Tl}$  and  $^{99\text{m}}\text{Tc}$  images in space and time, increased patient comfort, and higher clinical throughput [1]–[4]. However crosstalk, due to interactions in the body and imaging system as well as physical overlap of photopeaks with acquisition energy windows, degrades the quality of both the rest and stress images.

Effectively evaluating a three-class diagnostic task requires three-class receiver operating characteristic (ROC) analysis. To assess multiclass diagnostic performance, multiclass ROC analysis methods are required. However, development of these methods has remained an open theoretical problem since the introduction of binary ROC analysis in the 1950s [5], [6]. Motivated by the diagnostic problem of dual-isotope MPS imaging, we have previously developed a three-class ROC analysis methodology that extends and unifies the decision theoretic, linear discriminant analysis, and psychophysical foundations of binary ROC analysis in a three-class paradigm [7]–[10].

The goal of this work was to assess the image quality of dual-isotope MPS images and to perform a preliminary search for the optimal injected activity combination with respect to the performance of a three-class diagnostic task.

## I. BACKGROUND

### A. Dual-Isotope MPS Imaging

In a typical dual-isotope MPS study, two sets of scans will be acquired, a stress scan and a rest scan. In a typical dual-isotope protocol, a rest study is performed using  $^{201}\text{Tl}$  labeled radiopharmaceutical when the patient is at rest, and thus the resulting image characterizes the myocardial perfusion at rest.

Next, a stress study is performed using a  $^{99\text{m}}\text{Tc}$  labeled agent such as  $^{99\text{m}}\text{Tc}$  sestamibi injected after physical exercise or administration of a pharmacological stressor. The resulting stress image thus characterizes the status of myocardial perfusion under stress.

The stress and rest images are used together to identify three types of tissue: infarcted tissue where there is a defect, referred to as a fixed defect, in both the rest and stress images; ischemic tissue, having a compromised blood supply under stress, where there is a defect in the stress, but not the rest, image; normal tissue where there is no defect in either image. Note that knowledge of the defect status in one of the two images is only sufficient to make a partial diagnosis. The defect status of the rest image can be used to separate between a fixed defect and normal or a reversible defect; the defect status of the stress image can be used to separate between normal and a fixed or reversible defect. As a result, optimizing the performance on the three-class diagnostic task cannot be performed simply by separately optimizing the stress and rest images.

The diagnosis of cardiac disease using dual-isotope MPS requires both abnormality detection and characterization. To be specific, patients are sorted into one of the three categories as shown in Table I. In principle, the designation of class numbers is for the purpose of applying ROC analysis; i.e., any class can be assigned as Class 1, 2 or 3. However, in this particular problem, the class number assignments in Table I are more natural, as explained in the discussion of [7].

In one implementation of a simultaneous dual-isotope MPS acquisition protocol, the  $^{201}\text{Tl}$  labeled radiopharmaceutical is injected at rest and, subsequently,  $^{99\text{m}}\text{Tc}$  sestamibi is injected at peak stress. Different from the separate acquisition protocol, in which the rest scan is performed before the stress injection, both rest scan and stress scans are performed simultaneously immediately after the injection of  $^{99\text{m}}\text{Tc}$  under stress conditions. The stress ( $^{99\text{m}}\text{Tc}$ ) and rest ( $^{201}\text{Tl}$ ) images are simultaneously acquired in energy windows positioned over the  $^{201}\text{Tl}$  and  $^{99\text{m}}\text{Tc}$  photopeak energy windows. Photons emitted from  $^{99\text{m}}\text{Tc}$  decays can be scattered in the body or result in Pb x-ray fluorescence in the collimator and be detected in the  $^{201}\text{Tl}$  photopeak energy window. In addition, 135 keV and 167 keV  $^{201}\text{Tl}$  photons can be detected in the  $^{99\text{m}}\text{Tc}$  energy window. The combination of these two phenomena is referred to as crosstalk, and results in degradation of the two projection images.

Due to radiation dose concerns, the injected Tl activity is usually in the range of 2–4 mCi. On the other hand, the injected  $^{99\text{m}}\text{Tc}$  activity is typically in the range of 20–30 mCi depending on the size of the patient. This injected activity level was selected for dual-isotope MPS using sequential acquisition protocols to minimize the amount of crosstalk from the 135 and 167 keV  $^{201}\text{Tl}$  photons. For these injected activity levels, the crosstalk from Tl into the Tc energy window is so low that contamination of the rest image is a relatively minor problem. However, rigorous optimization of the injected activities is desirable to provide the best quality images for use in diagnosis.

As mentioned, one potential way to minimize the effects of crosstalk is to optimize the injected activity of the two isotopes. Since the quality of the  $^{201}\text{Tl}$  is typically poor, we will assume that it is kept fixed at the maximum allowed by radiation dose considerations. However, even with only a single parameter to vary, we are still faced with a dilemma: decreasing the  $^{99\text{m}}\text{Tc}$  injected activity (IA) decreases the crosstalk contamination of the rest ( $^{201}\text{Tl}$ ) image, and thus improves the rest ( $^{201}\text{Tl}$ ) image quality at the expense of noisier stress ( $^{99\text{m}}\text{Tc}$ ) images; similarly, increasing the  $^{99\text{m}}\text{Tc}$  injected activity improves the stress ( $^{99\text{m}}\text{Tc}$ ) image quality at the expenses of more crosstalk contamination on the rest ( $^{201}\text{Tl}$ ) images. In other words, we must consider the effect of injected activity on the quality of both images. Therefore, full optimization and evaluation of simultaneous dual-isotope MPS image quality requires analysis of a three-class diagnostic task, i.e., classifying the patient as normal or having a reversible or fixed defect.

## B. Binary ROC Analysis and Three-Class ROC Analysis

ROC analysis has solid theoretical foundations [11]–[19]. These foundations have resulted in the wide applicability of binary ROC analysis in the past, and will continue to ensure its prominence for the analysis of two-class problems. Much work has been done to extend binary to three-class [12], [20]–[41]. We believe that the most useful extensions of conventional binary ROC to multiclass ROC will be based on extensions of these theoretical foundations.

Motivated by the medical diagnostic problem of dual-isotope MPS, we have developed a three-class ROC analysis method [7]–[10]. The proposed three-class methodology has been proved to extend and unify the decision theoretic [7], [8], linear discriminant analysis [9], and psychophysical foundations [42] of conventional binary ROC analysis in a three-class paradigm.

This three-class ROC analysis methodology has a decision theoretic foundation: it provides decision variables and decision rules that maximize the expected utility by assuming that incorrect decisions have equal utilities under the same hypothesis (the equal error utility assumption) [7]. The proposed three-class decision model uses a pair of log likelihood ratios

as the optimal decision variables that span the decision space, as shown in Fig. 1(a). Sweeping the structure represented by the bold lines, termed the decision structure, across the decision plane, and calculating the fractions of correct decisions under each hypothesis maps out an ROC surface in the 3-D space spanned by the fractions of correct decisions in each class, i.e.,  $T1F$ ,  $T2F$ , and  $T3F$  Fig. 1(b). Here,  $T_iF$  represents the fraction of correct decisions on class  $i$  when the truth is class  $i$ , and is termed the sensitivity for this class. We have proved that the volume under three-class ROC surface (VUS) is a figure-of-merit for three-class task performance [7].

We then investigated the proposed three-class decision model with respect to the decision theoretic foundation [8]. In particular, we found that the proposed decision model 1) maximizes the expected utility (MEU) under equal error utility assumption, 2) maximizes the probability of making correct decisions according to (maximum correctness MC) criterion, 3) satisfies the Neyman-Pearson (N-P) criterion in a sense that it maximizes the sensitivity of one class given the sensitivities of the other two, and 4) the resulting three-class ROC surface contains the maximum-likelihood (ML) operating point.

In order to further enhance the applicability of the proposed three-class ROC analysis method, we developed a three-class mathematical observer, the three-class Hotelling observer (3-HO), that computes optimal linear observer test statistics from a set of input images [9]. The 3-HO uses only the first- and second-order ensemble statistics and maximizes the signal-to-noise ratio (SNR) of the test statistics between each pair of the classes simultaneously under certain data normality conditions (to be explained later).

We subsequently illuminated the psychophysical foundation of three-class ROC by proving that VUS equals the percent correct (PC) in a three-class categorization procedure, i.e.,  $VUS = PC$  [42]. In the three-class categorization procedure, three objects randomly sampled from each of the three distinct classes are presented to the observer at the same time. The observer is instructed to categorize all three objects into the corresponding hypothesis. The observer is said to make a correct decision when and only when all three objects are correctly categorized. The relationship  $VUS = PC$  is true regardless of the decision variables used or their distributions under different hypotheses. This relationship was first proved by Scurfield [39]. In previous work, we proved this relationship using a different approach and carefully pointed out that this relationship holds for continuous data; it has not been proven that this relationship can be extended to discrete data [42]. Although this relationship does not depend on the decision variables used, it does depend on the decision structure (defined by the decision rules) used. Currently this relationship has only been proved for the decision structure shown in Fig. 1. This relationship is analogous to the equivalence of the AUC and percent correct on a two-alternative forced choice task in the binary case. In addition to providing psychophysical meaning for the VUS, this property provides an alternative method for calculating the VUS which is both more computationally efficient and more amenable to statistical analysis [42].

## II. Application to Task-Based Image Quality Assessment

As the first medical imaging application of the proposed three-class ROC analysis methodology, this study seeks the optimal injected activity combination in simultaneous dual-isotope MPS imaging and provides an example for future medical applications of this methodology. The sections will be organized to parallel the four key elements described in [13] to provide a step-by-step description of the application of this methodology to task-based image quality assessment to seek the optimal injected activity.

## A. Identification of the Task

We have described the diagnostic problem using dual-isotope MPS technique in the Background section. Table I shows the definition of the three classes.

## B. Knowledge of the Objects to be Imaged and the Imaging Process-MPS Data Generation

As the first rigorous medical application of the proposed three-class ROC analysis method, we chose to use simulated images and mathematical observers for the following reasons. First, computer simulations allow us to know exactly the true state of the simulated data, which is essential in both binary and three-class ROC studies. Second, simulations can provide a large number of images that can be used as inputs for mathematical observers. A mathematical observer study has the advantage that it can provide figures-of-merit for many more combinations of parameters than is practical (or possible) with human observer studies. Finally, the large number of images and lack of inter-observer variability will provide smaller confidence intervals and thus make it easier to rank methods and parameters.

Having identified the task, in this study we generated, as described below, a population of stress ( $^{99m}\text{Tc}$ ) and rest ( $^{201}\text{Tl}$ ) images that realistically model the patient variations and the imaging process.

**1) Phantom Population Generation**—We used a previously developed phantom population [43], [44] based on the 4-D NCAT phantom [45] that realistically models anatomical and organ uptake variations seen in clinical data. The phantom population includes 24 NCAT anatomies with half male and half female. Anatomical variations in the patient size, heart angle, size, and position, and lung volume were modeled based on those observed in a patient population [46]. In addition, we modeled variations in perfusion defect sizes and locations. The details of the phantom population generation are described in [44] and [47]. The mechanism for modeling and producing variation in organ activity uptake are described below.

**2) Projection Population Generation**—For each of the 24 anatomies, we separately simulated projections of five organs (heart, lung, body, liver and kidney) [48]. These projections were generated using a combined SimSET and angular response function (ARF) simulation code [49], [50]. The ARF code provides an efficient way to model interactions in the collimator such as Pb X-rays, an important source of crosstalk in  $^{99m}\text{Tc}/^{201}\text{Tl}$  dual-isotope MPS imaging. For each organ of each anatomy, the  $^{99m}\text{Tc}$  projection was simulated using a 140keV photon energy and a 20% wide energy window centered on 140 keV; the  $^{201}\text{Tl}$  energy window was 20% wide and centered at 70 KeV. We simulated seven major  $^{201}\text{Tl}$  emission energies: 70.8, 68.9, 80.3, 167.4, 79.8, 82.6, and 135.0 KeV based on the corresponding abundances. The effects modeled include attenuation, scatter in the body, detector response, penetration and scatter in the collimator, and crosstalk contamination in the  $^{99m}\text{Tc}$  and  $^{201}\text{Tl}$  energy windows.

To generate a population of projection data that includes realistic variations in anatomy and organ uptake, we then scaled and summed the organ projections for each anatomy to create projections with variations in uptake ratios, count level, and defect contrast. The scaling factors were used such that the resulting uptake ratios and count levels of  $^{99m}\text{Tc}$  and  $^{201}\text{Tl}$  images were comparable to those in a population of 34 patients who underwent  $^{99m}\text{Tc}$  stress imaging and 50 patients who underwent  $^{201}\text{Tl}$  rest imaging, respectively. Defects at five locations in the myocardium were simulated. The projection data were scaled to model three different  $^{99m}\text{Tc}:^{201}\text{Tl}$  injected activity combinations: 24:3 mCi, 16:3mCi, and 8:3 mCi. Projection data modeling a simultaneous acquisition protocol was generated by summing

crosstalk from the simulation using one isotope into the corresponding energy window for the other isotope. Triplets of projection data modeling the three diagnostic classes were created by subtracting appropriately scaled defect projections from the appropriate set of no-defect projection data. For each injected activity combination, we thus had 4320 pairs of  $^{99m}\text{Tc}/^{201}\text{Tl}$  projection data (24 phantoms  $\times$  5 defect locations  $\times$  6 uptake ratio combinations  $\times$  3 defect statuses, i.e., normal, reversible defect, and fixed defect  $\times$  2 acquisition protocols).

**3) Reconstruction Generation and Post-Reconstruction Processing**—We reconstructed the 4320 pairs of  $^{99m}\text{Tc}/^{201}\text{Tl}$  images using the OS-EM reconstruction algorithm with compensation for attenuation, detector response, and scatter (ADS) without compensation for crosstalk contamination. For each stress ( $^{99m}\text{Tc}$ ) image, we used 15 iterations with 16 updates per iteration; for each rest ( $^{201}\text{Tl}$ ) image, we used 15 iterations with four updates per iteration. Reconstructed images for all iterations were saved for use in finding the optimal number of iterations as described below.

The reconstructed images were then processed using standard post-processing methods including reorientation, Butterworth filtering, and windowing to obtain short axis (SA)  $^{99m}\text{Tc}/^{201}\text{Tl}$  image pairs. For each reoriented 3-D stress ( $^{99m}\text{Tc}$ ) image, a  $32 \times 32$  image centered on the heart and located at the centroid of the defect (or the same short-axis slice in the corresponding defect-absent image) was extracted from the postprocessed short-axis slices. The same short-axis slice in the corresponding rest ( $^{201}\text{Tl}$ ) image was also selected. These  $^{99m}\text{Tc}/^{201}\text{Tl}$  SA image pairs served as the input for the three-class mathematical observer studies. Fig. 2 shows an example of a triplet of dual-isotope MPS objects.

### C. Identification and Knowledge of the Observer That Will Perform the Task

Having identified the task and described the generation of the data, we will now describe the observer used.

In [9], we developed a three-class channelized Hotelling observer (3-HO) and investigated its properties. We showed that when the pairwise binary Hotelling templates satisfy the linear relationship

$$\vec{w}_{23} - \vec{w}_{13} = \vec{w}_{21} \quad (1)$$

where  $\vec{w}_{ij}$  is the binary Hotelling template between classes  $i$  and  $j$ , the 3-HO maximizes the SNR of the test statistics between each pair of the classes. Furthermore, when the data follow a Gaussian distribution with equal covariance, the 3-HO estimates the three-class ideal observer (3-IO) performance and maximizes the VUS.

In this study, the goal was to develop and apply an observer, the 3-CHO, which has the potential to model human performance in a three-class task. This was done based on the large body of previous experience in modeling human observer performance using the binary CHO, e.g., [44], [47], [51]. In addition to its potential for modeling human observer performance, the 3-CHO eliminates the computational difficulties that arise in applying the 3-HO directly to the SA images. These difficulties arise due to the requirement to estimate the covariance matrix which, for the  $2048 \times 1$  data vector containing the two  $32 \times 32$  SA images (stress and the rest), results in a  $2048 \times 2048$  covariance matrix. Estimating the 3-HO requires inverting this matrix, which, though computationally tractable, requires precise estimation of the matrix, and thus a large ensemble of data vectors. In order to obtain a covariance matrix that is invertible, a minimum of  $2048+1$  data vectors are required in the

training ensemble. In this 3-CHO study, the SA images were first processed using appropriate frequency domain channels. The channelized data vector ensemble, sometimes referred to as feature vector ensemble, was then used to train and test the CHO just as the original data vectors would have been used for the HO. The frequency domain channels used have produced good correlation with human observer performance in previous binary CHO studies [51], [52], and provide data reduction that drastically reduces the size of the data vector ensemble needed to estimate the 3-HO.

### 1) Preparing the Feature Vector Population to be Processed, or “Observed” by the 3-CHO—

We are interested in comparing the image quality of  $^{99m}\text{Tc}/^{201}\text{Tl}$  rest/stress images for different injected activity combinations. To make a meaningful comparison, we grouped images according to their data treatments. Each data treatment refers to a specific combination of injected activities, acquisition protocols, reconstruction algorithms, iteration numbers, and postprocessing filter cutoff frequencies for both  $^{99m}\text{Tc}/^{201}\text{Tl}$  images. Thus, a three-class dataset for each data treatment has 360 (24 phantoms  $\times$  5 defects  $\times$  6 uptake ratio combinations) triplets of  $^{99m}\text{Tc}/^{201}\text{Tl}$  rest/stress image pairs representing the images from normal patients and patients with reversible and fixed defects.

For each  $^{99m}\text{Tc}/^{201}\text{Tl}$  image pair, we denote the stress ( $^{99m}\text{Tc}$ ) short-axis image vector as  $\vec{\mathbf{g}}_s$  and the rest SA image vector as  $\vec{\mathbf{g}}_r$ . Both  $\vec{\mathbf{g}}_s$  and  $\vec{\mathbf{g}}_r$  are  $N \times 1$  vectors, where  $N$  equals the total number of pixels on each image, i.e.,  $TV = 32 \times 32 = 1024$ . We stack the two image vectors together to form a  $2N \times 1$  image vector  $\vec{\mathbf{g}}$ , i.e.,

$$\vec{\mathbf{g}} = \begin{bmatrix} \vec{\mathbf{g}}_s \\ \vec{\mathbf{g}}_r \end{bmatrix}. \quad (2)$$

Thus, a three-class dataset for each data treatment has 360 triplets of image vectors representing the image pairs from normal patients, and patients with reversible and fixed defects.

$$\vec{\mathbf{v}} = \begin{bmatrix} \vec{\mathbf{v}}_s \\ \vec{\mathbf{v}}_r \end{bmatrix} = \begin{bmatrix} \vec{\mathbf{u}}_1^\dagger \vec{\mathbf{g}}_s & \vec{\mathbf{u}}_2^\dagger \vec{\mathbf{g}}_s & \dots & \vec{\mathbf{u}}_J^\dagger \vec{\mathbf{g}}_s & \vec{\mathbf{u}}_1^\dagger \vec{\mathbf{g}}_r & \vec{\mathbf{u}}_2^\dagger \vec{\mathbf{g}}_r & \dots & \vec{\mathbf{u}}_J^\dagger \vec{\mathbf{g}}_r \end{bmatrix}^t \quad (3)$$

Since both stress and rest images were presented to the observer simultaneously, based on our knowledge of binary defect-detection task [44], [47], [51], we propose the following channel mechanism. First, the stress image and rest image of each patient are processed using appropriate set of channels  $U = [\vec{\mathbf{u}}_1 \vec{\mathbf{u}}_2 \dots \vec{\mathbf{u}}_J]$ , respectively, where  $J$  is the number of channels. The feature vectors for both stress and rest images,  $\vec{\mathbf{v}}_s$  and  $\vec{\mathbf{v}}_r$ , are then stacked together to form a feature vector  $\vec{\mathbf{v}}$  for data vector  $\vec{\mathbf{g}}$ . Using these, the observer study is performed using the feature vector  $\vec{\mathbf{v}}$  instead of the data vector  $\vec{\mathbf{g}}$ .

This process can be expressed as shown in (3) at the bottom of the page, which is equivalent to applying the channel matrix

$$\Psi = \begin{bmatrix} \vec{\mathbf{u}}_1 & \vec{\mathbf{u}}_2 & \dots & \vec{\mathbf{u}}_J & \vec{\mathbf{0}} & \vec{\mathbf{0}} & \dots & \vec{\mathbf{0}} \\ \vec{\mathbf{0}} & \vec{\mathbf{0}} & \dots & \vec{\mathbf{0}} & \vec{\mathbf{u}}_1 & \vec{\mathbf{u}}_2 & \dots & \vec{\mathbf{u}}_J \end{bmatrix} \quad (4)$$

to the data vector, where  $\vec{\mathbf{0}}$  is a zero column vector of  $N$  elements and  $N$  is the number of pixels in the channel image, which is equal to the number of pixels in each SA image. The feature vector  $\vec{\mathbf{v}}$  is given by

$$\vec{\mathbf{v}} = \Psi^{\dagger} \vec{\mathbf{g}}. \quad (5)$$

In the present study, four octave-wide rotationally symmetric frequency channels, described in detail in [44], [47], [51], were used as the channels  $U$ . After applying (4), for each data treatment, there were 360 triplets of feature vectors representing the rest/stress image pairs from normal patients, and patients with reversible and fixed defects.

**2) Properties of the 3-CHO**—As described above, a 3-CHO can be treated as a 3-HO applied to channelized data (feature) vector ensemble. Thus, (1) must be satisfied for the 3-CHO to maximize the SNR of the test statistics between each pair of the classes simultaneously, where  $\vec{\mathbf{w}}_{ij}$  is the binary Hotelling template between classes  $i$  and  $j$  computed using the feature vector ensemble. Furthermore, when the feature vectors follow Gaussian distributions with equal covariance, the 3-CHO estimates the log likelihood ratios and maximizes the VUS. In the following, we describe the procedures used to check whether the statistical properties of the three-class dual-isotope MPS feature vector ensemble meets these conditions.

**a) Checking the linear property of the three-class channelized dataset:** We first tested whether the data satisfy (1). A “three-template test” was developed to test the linearity. In this test,  $\vec{\mathbf{w}}_{13}$ ,  $\vec{\mathbf{w}}_{23}$ , and  $\vec{\mathbf{w}}_{21}$  were first computed using the following:

$$\vec{\mathbf{w}}_{13} = \left( \frac{1}{2} K_1 + \frac{1}{2} K_3 \right)^{-1} (\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_3) \quad (6)$$

$$\vec{\mathbf{w}}_{23} = \left( \frac{1}{2} K_2 + \frac{1}{2} K_3 \right)^{-1} (\vec{\mathbf{v}}_2 - \vec{\mathbf{v}}_3) \quad (7)$$

and

$$\vec{\mathbf{w}}_{12} = \left( \frac{1}{2} K_1 + \frac{1}{2} K_2 \right)^{-1} (\vec{\mathbf{v}}_1 - \vec{\mathbf{v}}_2) \quad (8)$$

where  $\vec{\mathbf{v}}_i$  is the mean feature vector of class  $i$  and  $K_i$  is the covariance matrix of  $i$ th class feature vectors. In a 3-D space, we plotted all triplets  $(x, y, z) = (\vec{\mathbf{w}}_{23}^t \vec{\mathbf{v}}, \vec{\mathbf{w}}_{13}^t \vec{\mathbf{v}}, \vec{\mathbf{w}}_{21}^t \vec{\mathbf{v}})$ , which should be on the plane  $x - y = z$ , provided (1) holds. We fit the triplets  $\{(\vec{\mathbf{w}}_{23}^t \vec{\mathbf{v}}, \vec{\mathbf{w}}_{13}^t \vec{\mathbf{v}}, \vec{\mathbf{w}}_{21}^t \vec{\mathbf{v}})\}$  with a plane, and the error between the fitted plane and plane  $x - y = z$  tells how well (1) holds.

**b) Checking the normality of the three-class feature-vector population:** To test whether the two test statistics obtained using the two 3-CHO templates estimate the log likelihood ratios, the normality of the set of feature vector  $\{\vec{\mathbf{v}}\}$  was investigated as well as the equivalence of the covariance matrixes. It is very hard to study the multivariate normality of  $\{\vec{\mathbf{v}}\}$  directly, while the statistical properties of the test statistics can be easily studied. The test statistics are a linear transformation of the feature vector  $\vec{\mathbf{v}}$  and should be Gaussian

provided the  $\vec{v}$  follow a multivariate Gaussian distribution. The normality of the test statistics is thus a necessary but insufficient to prove the normality of the feature vector.

**c) Compare the covariance matrixes of the three classes:** To assess the equivalence of the covariance matrixes of the three classes, we simply compared them in terms of profiles of the diagonal elements.

**3) Performing the 3-CHO Study**—After generating the three-class feature vector ensemble for each data treatment, we used half the population to train the 3-CHO. In this training procedure, two three-class Hotelling templates used were  $\vec{w}_{13}$  and  $\vec{w}_{23}$  as expressed in (6) and (7), respectively. The 3-CHO was then tested using the remaining half of the feature vector ensemble to produce an ensemble of test statistic pairs.

#### D. Figure-of-Merit Provided by Three-Class ROC Analysis

Having obtained the 3-CHO decision plane using the feature vectors from each data treatment, we applied the proposed three-class ROC analysis method and obtained the volume under surface (VUS) value for each data treatment using the approach described in [10]. This VUS value served as the figure-of-merit to quantify the task performance for the corresponding data treatment.

### III. Experiments and Results

The major purpose of the present paper was to develop a framework for three-class task-based dual-isotope MPS image quality assessment using simulations studies. Thus, instead of performing a comprehensive evaluation and optimization over all possible data treatments, we investigated a few data treatments and will describe the results from two experiments below. The first experiment was 3-CHO study for a single data treatment. In this experiment, we focused on application of a 3-CHO and demonstration of the resulting three-class ROC decision space as well as the ROC surface. The second experiment compared task performance as measured by the 3-CHO and binary CHO performance for different  $^{99m}\text{Tc}/^{201}\text{Tl}$  injected activity combinations. The purpose of the second experiment was to demonstrate the three-class methodology is essential for optimizing the injected  $^{99m}\text{Tc}$  in dual-isotope  $^{99m}\text{Tc}/^{201}\text{Tl}$  MPS.

#### A. 3-CHO Study

In this section, we show results from a 3-CHO study using the data treatment whose parameters are shown in Table II. The parameters shown might not be globally optimal. However, they were chosen based on our previous knowledge of the optimal parameters for binary observer performance.

**1) Results of the Three-Template Test**—Fig. 3 shows the result of the three-template test for this data treatment. After plotting  $(\vec{w}_{23}^t \vec{v}, \vec{w}_{13}^t \vec{v}, \vec{w}_{21}^t \vec{v})$  for each feature vector in the 3-D space, we rotated the space three dimensionally. In a certain orientation, all points align in one straight line as shown in Fig. 3(a), indicating all the points are located on one 2-D plane in the 3-D space; Fig. 3(b) shows only 10 triples of data for a clearer view. This demonstrates the linear dependence of the three templates, as described in (1).

To quantify how well the experimental plane fits the theoretical plane  $x - y = z$ , we then fitted the triplets of  $(x, y, z) = (\vec{w}_{23}^t \vec{v}, \vec{w}_{13}^t \vec{v}, \vec{w}_{21}^t \vec{v})$  with a plane of the form  $x + by + cz + d = 0$  using linear least square fitting techniques. We also calculated the root mean square residual error (RSMRE) between the fitted plane and the ideal plane. This error is defined as

$$\text{RMSRE} = \sqrt{\frac{\sum_i^N ((x_i - y_i) - z_i)^2}{L}} \quad (9)$$

where  $L$  is the total number of triplets. These results demonstrate how well the estimated templates satisfy (1). In particular, Table III demonstrates that the root mean square residual error is 0.2769 when we pool the triplets from all three classes together. This means that, for each triplet, the mean square error in the  $z$  direction is 0.2769. Compared with the data range in  $z$  direction, i.e., the maximum and the minimum of  $z_i$ , which are 0.8141 and 8.9331, the error 0.2769 is relatively small ( $< 0.3$ , or  $< 3.5\%$  of the data range) and the parameters defining the plane are close to ideal values. We thus concluded that these data satisfy (1) very well.

**2) Statistical Properties of the Simulated Dual-Isotope Data**—Next, we tested the normality and compared the covariance matrixes of the test statistics of the data to see whether the 3-CHO estimates the log LRs. Figs. 4–6 show the histograms of the two-class test statistics for each pair of the classes. We then tested whether these distributions are consistent with Gaussians distributions. To do so, we used the Jarque-Bera test to check the normality of each of the test statistic distributions. The null hypothesis for these tests was that the data were normally distributed with unspecified mean and variance; the alternative hypothesis was that the data were not normally distributed. We found that the binary Hotelling test statistics in every pairwise comparison follow normal distribution at a level of  $P = 0.01$ , as demonstrated by Tables IV–VI.

Fig. 7 shows comparisons of the covariance matrixes of the three classes. A profile is drawn along the diagonal on the three covariance matrixes. As can be seen, the covariance matrix for data from the normal class has a bigger magnitude than that of the other two classes at the fourth element. However, the overall trends of the magnitude of the diagonal elements are very similar among the three classes.

**3) Decision Plane and ROC Surface of Simulated Data**—We have shown that the data satisfy (1) reasonably well, and therefore the 3-CHO can be applied to these data to maximize the SNRs of each pair of the classes simultaneously. The normality and equivalence tests suggest Gaussian distributed data with similar covariance matrixes, thus we conclude that the 3-CHO provides a reasonable estimate to the log likelihood ratios. However, since the agreement of the three-template test and covariance matrixes is not exact and has no tests of statistical significance have been performed, it is difficult to quantify how well the 3-CHO estimates the log likelihood ratios.

By applying the three-class decision model and 3-CHO to the simulated dual-isotope data, we obtained the decision planes and the corresponding ROC surface. Fig. 8 shows the decision plane with a scatter plot of the pairs of test statistics for each class; Fig. 9 shows the corresponding three-class ROC surface. We fitted the test statistic distributions of each class with bivariate Gaussian distribution, and then mapped out a three-class ROC surface based on the fitted distributions, as shown in Fig. 10. The volume under surface value (VUS) of the fitted surface is 0.704, suggests good classification performance in comparison with the degenerate (pure-guessing) VUS of  $1/6$  (0.167).

## B. Application of Three-Class Methodology to the Preliminary Evaluation and Optimization of Simultaneous Dual-isotope MPS

In the previous sections, we have described a 3-CHO study that produced a VUS value for a particular data treatment. To demonstrate the three-class methodology is essential for

optimizing the injected  $^{99m}\text{Tc}$  in dual-isotope  $^{99m}\text{Tc}/^{201}\text{Tl}$  MPS, in this section, we will compare task performance as measured by the 3-CHO and binary CHO performance for different  $^{99m}\text{Tc}/^{201}\text{Tl}$  injected activity combinations. The data treatments in Fig. 7 will be used for both simultaneous acquisition and separate acquisition.

**1) Separate Acquisition**—We first studied the case of separate acquisition of the stress and rest images, i.e., where there is no crosstalk between the two images.

We performed two kinds of binary observer studies. In the first study, we pooled the defect-absent stress ( $^{99m}\text{Tc}$ ) images from normal class and defect-present stress images from the reversible classes together, and then investigated the image quality of the stress images; similar investigations were performed for the rest ( $^{201}\text{Tl}$ ) images using the defect-absent rest images from the normal classes, and defect-present rest images from the fixed defect class. The results are shown in Fig. 11. Increasing the  $^{99m}\text{Tc}$  injected activity while keeping the  $^{201}\text{Tl}$  injected activity unchanged results in an improvement of stress ( $^{99m}\text{Tc}$ ) image quality, as measured by the AUC value obtained from the stress images. The rest ( $^{201}\text{Tl}$ ) image quality, on the other hand, remains unchanged. As a result, we see the overall image quality, as measured by the VUS value, improves with the increase of  $^{99m}\text{Tc}$  injected activity, just as one would expect for a separate acquisition protocol.

The second binary observer study involved pairwise comparisons between each pair of classes, i.e., the AUC values for normal versus reversible defect classes, reversible versus fixed defect classes, and normal versus fixed defect classes. This is because we are considering the stress/rest imaging, the rest and stress images are usually interpreted simultaneously. Fig. 12 shows a comparison of pairwise binary observer performance with three-class observer performance for separate acquisition. As the  $^{99m}\text{Tc}$  injected activity increases, the quality of the ( $^{99m}\text{Tc}$ ) stress image quality should improve due to reduced noise, and the  $^{201}\text{Tl}$  rest image quality should remain unchanged. We can see this behavior in the pairwise three-class data. For these data, the only pair of classes where the rest image has a defect in one class and no defect in the other class is for patients with reversible versus fixed defects. In this case, the change in noise in the stress image should have no impact on the ability to separate these classes. This is just what is observed: the AUC does not change as the  $^{99m}\text{Tc}$  injected activity is increased. On the other hand, there are two pairs of classes where the stress image has a defect in one class and no defect in the other class: normal versus fixed and normal versus reversible. In the latter case, there is no change in the defect status in the rest image, so we might expect the reduction in noise to have a more pronounced effect on the ability to separate these two classes. However, even for normal versus fixed, the reduced noise in the stress image should make it easier to detect the stress defect and thus easier to perform the classification. Again, this is just what is observed in Fig. 12. For the full three-class task we would expect that improving the stress image quality should improve performance on the three-class task. Again, this is what is observed in Fig. 12. Thus, the changes in VUS and AUC make intuitive sense for the case of separate acquisition.

**2) Simultaneous Acquisition**—We now turn to the case of simultaneous acquisition. In this case, we expect that increasing the injected  $^{99m}\text{Tc}$  activity will improve stress image quality. However, it will increase the crosstalk in the rest image and thus degrade rest image quality. It is reasonable to expect that there will be an optimal injected activity. Similar to the separate acquisition, we performed two binary observer studies, and compared the results with the three-class observer study.

The first binary observer study uses pooled defect images, just as described for the separate acquisition case. The results are shown in Fig. 13. Increasing the  $^{99m}\text{Tc}$  injected activity

while keeping the  $^{201}\text{Tl}$  injected activity unchanged results in an improvement of stress ( $^{99\text{m}}\text{Tc}$ ) image quality, as measured by the AUC value of the stress images. The rest ( $^{201}\text{Tl}$ ) image quality, however, was degraded due to the increased crosstalk. Thus, the binary observer studies using the stress and rest images do not provide a way to optimize or select an appropriate injected activity combination. However, the VUS value obtained from three-class ROC analysis provides evidence of the existence of an optimum.

Fig. 14 shows a comparison of pairwise binary observer performance with three-class observer performance for simultaneous acquisition. For the case of the pairwise binary classification, performance for reversible versus fixed degrades with increasing  $^{99\text{m}}\text{Tc}$  image quality, as the increased crosstalk makes it more difficult to see the defect to the rest image. However, the ability to distinguish between normal and reversible defects improves. This is because the reduced stress image noise makes it easier to see the stress defect; the crosstalk in the rest image is irrelevant since there is no difference in rest defect status between these two classes. However, for the case of normal versus fixed, there seem to be an optimal injected activity. The change of VUS value with the injected activities also suggests there is an optimal injected activity. Thus, we see that, using only pair-wise binary classification performance, we are unable to unambiguously find the optimal injected activity as the data from the pairwise binary tasks provides contradictory information. On the other hand, the three-class data provides evidence of the existence of an optimum and, with finer sampling of the injected  $^{99\text{m}}\text{Tc}$  activity, could be used to locate the maximum.

Both Figs. 13 and 14 show that with the VUS only varies slightly with an increase of injected  $^{99\text{m}}\text{Tc}$  activity, in contrast to the larger changes in AUC values. This suggests that the three different injected activity combinations provide similar image quality. Given the fact that  $24\text{mCi}^{99\text{m}}\text{Tc}$  is used clinically, this study suggests that the patient  $^{99\text{m}}\text{Tc}$  injected activity can be significantly reduced without jeopardizing the overall image quality.

As shown in both Figs. 13 and 14, the error bars on the VUS are relatively large compared with the differences between the VUS values. This is due to the inclusion of anatomical and uptake variations in the projection population, which are included to increase the clinical relevance, and the relatively small number of cases in each class. More cases, as well as more injected activities, are needed in a future study to more definitively locate the optimal injected activity. In addition, more sophisticated methods for performing hypothesis testing that model case sample correlations are needed to investigate the statistical power of three-class ROC methodology.

## IV. Discussion

### A. Equal Error Utility Assumption

Our proposed three-class ROC analysis method was derived based on decision theory by maximizing the expected decision utility and assuming that incorrect decisions have equal utilities under the same hypothesis (equal error utility assumption). Consequently, one might wonder whether three-class ROC analysis is applicable to dual-isotope MPS since incorrect decisions might have quite different utilities (or consequences) under the same hypothesis. For example, when the patient has a fixed defect, mistaking him or her for normal will result in no therapeutic interventions, while mistaking him or her for having a reversible defect might potentially result in referral for a needless and ineffective revascularization procedure. Apparently, the consequences, and thus the utilities, of the two types of false decisions given the same hypothesis are not equivalent. As a result, the usefulness of the proposed three-class ROC methodology for this application is brought into question.

However, as described in the previous section, we have demonstrated the optimality of the proposed three-class ROC methodology with respect to four decision criteria, namely MEU, MC, N-P, and ML criteria [8]. In this case, we appeal to the N-P criterion to justify the use of the proposed three-class ROC analysis method for this particular medical application [9]. Although N-P applied to the sensitivities and MEU with the equal error utility assumption result in the same three-class ROC surface, they are two distinct decision criteria and it is difficult to relate the set of parameters that define a given operating point on the ROC surface. In other words, each operating point on the ROC surface has an N-P explanation. No knowledge of utilities or prevalences is needed to find this operating point, though there is an implicit set of utilities and prevalences that are defined for each such operating point. The argument provided below is not unique to the three-class ROC research, but very similar to the argument between the Bayesian and Frequentist schools of thoughts in decision making.

In particular, in the diagnosis of coronary artery disease, the doctor might want to be very sensitive to finding both fixed and reversible defects. This is based on the doctor's general knowledge about the utilities and prevalence for these classes and this knowledge can be quite incomplete, which is most of the case in clinical practice. Given this prior knowledge, the doctor might specify two desired sensitivities for detecting reversible and fixed defects. It is likely easier for a doctor to specify the two desired sensitivities than to assign numerical utilities to all the possible decision outcomes. The optimal thing to do in terms of maximizing the sensitivity for normal patients given the sensitivities for fixed and reversible defects is to use the proposed three-class decision model. The three-class ROC surface resulting from application of the model describes the maximum sensitivity for normal patients for any such operating point described in terms of sensitivities for fixed and reversible defects.

While the proposed ROC methodology might not maximize performance for all combinations of utilities and prevalences, in practice, it is unlikely that it would be possible to know the exact prevalences and utilities for all decisions. It is more reasonable to believe that a physician could specify a set of acceptable sensitivities for two of the classes based on his incomplete knowledge of utilities and disease prevalences. Thus, we argue that optimizing the injected activity using the proposed three-class ROC methodology is likely to provide operating points that will sufficiently optimal for clinical practice.

As a task-based assessment method, ROC analysis summarizes the classification of a particular observer on a particular clinical task. Since the observers, e.g., doctors or computer aided diagnosis software, do not have access to the complete and exact information about the prior parameters required for MEU, they cannot apply the MEU criterion in practice. In other words, "even if we analysts could handle that many numbers, we could hardly expect future image readers to adjust so large a number of response biases to near-optimal settings. Even computer-aided diagnostic systems, would experience difficulty with that many quantities" [19].

Despite the above reasons that the proposed three-class methodology might prove adequate for optimizing clinical tasks, this remains to be demonstrated. Ultimately, this method should be compared and contrasted with a fully general three-class ROC analysis methodology, assuming there is one exists. However, in the meantime, the proposed methodology is arguably the most theoretically rigorous and relevant available, and its use is thus justifiable for the proposed application.

## B. Limitations and Future Study

We have proposed a three-class ROC analysis that, although not fully general in a maximum expected utility sense, is general in a sense that it extends and unifies decision theoretic, linear discriminant analysis and psychophysical foundations of binary ROC analysis in a three-class paradigm. This pilot study demonstrates the feasibility of applying this method to a clinically relevant problem.

As the first medical application of the proposed three-class ROC analysis, this study has several limitations, which will be addressed in future research. First, partly due to the lack of sophisticated statistical methods for three-class ROC analysis, the statistical analysis in this pilot study is not ideal. Developing statistical analysis methods for significance testing, power prediction and quantifying the linearity of the data statistics are areas that deserve future attention. Second, the 3-CHO implemented in this work is a natural extension of a binary CHO that has been shown to predict human observer performance in MPS images. However, there is no guarantee that this correlation will hold for the three-class task investigated here. Future research thus involves validating this model observer as well as the corresponding data analysis methods. Third, as mentioned before, the error bars on the VUS values in both Figs. 13 and 14 are relatively large compared with the differences between the VUS values. Thus a larger number of cases will be employed in future follow-up studies that aim to more definitively and precisely find the optimal injected activity combination. Finally, as future research, we will apply the proposed three-class ROC analysis to comprehensively evaluate and optimize the various compensation methods used in dual-isotope MPS.

## V. Conclusion

This paper presents the first medical application of a previously-proposed three-class ROC analysis method. We have presented a step-by-step description of a set of methods that are critical to conducting a three-class ROC study. We compared results of binary pairwise binary ROC studies with more compressive three-class study, and demonstrated the necessity of three-class ROC in evaluating a three-class diagnostic task.

In particular, we applied this method to the problem of optimizing  $^{99m}\text{Tc}$  injected activity in simultaneous dual-isotope MPS imaging. We have shown that binary ROC analysis is not enough for the evaluation and optimization of simultaneous dual-isotope MPS imaging. Considering the three-class nature of simultaneous MPS imaging provides a more general framework for evaluation of MPS imaging systems and algorithms, and, in some cases, is essential for its evaluation and optimization.

The data suggest an optimal injected activity of between 8 and  $24\text{mCi}^{99m}\text{Tc}$  for the stress study with a fixed  $3\text{mCi}^{201}\text{Tl}$  injected activity for the rest study. However, further investigation is warranted to find this optimum more accurately and precisely. In addition, the change in VUS values was not large as the  $^{99m}\text{Tc}$  activity was increased from 8 and 24 mCi in simultaneous acquisition. This suggests that there is the potential to reduce the injected  $^{99m}\text{Tc}$  without jeopardizing the overall image quality.

Finally, this paper provides a practical example of applying this three-class ROC analysis method to medical problems. We believe that the availability of these methods and a practical example of how to use them will elicit more medical applications of multiclass ROC analysis and more methodological investigations of the multiclass classification problem.

## Acknowledgments

This work was supported by the National Institutes of Health under Grant K99-EB007620, R01-EB000288 and Grant R01-HL068575. The content of this work is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health or its various institutes.

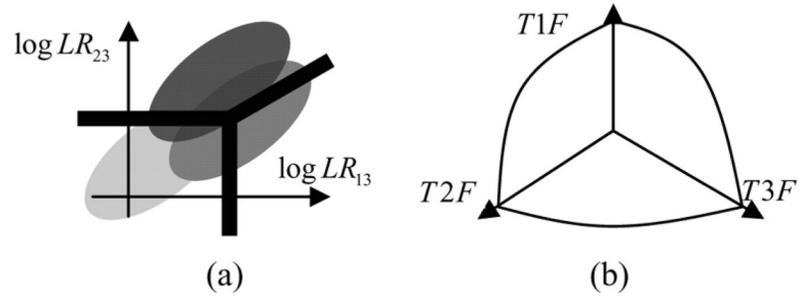
The authors would like to thank their colleagues Dr. J. M. Links and Dr. B. M. W. Tsui, for the helpful comments and thought-provoking discussions with regard to this work and manuscript.

## References

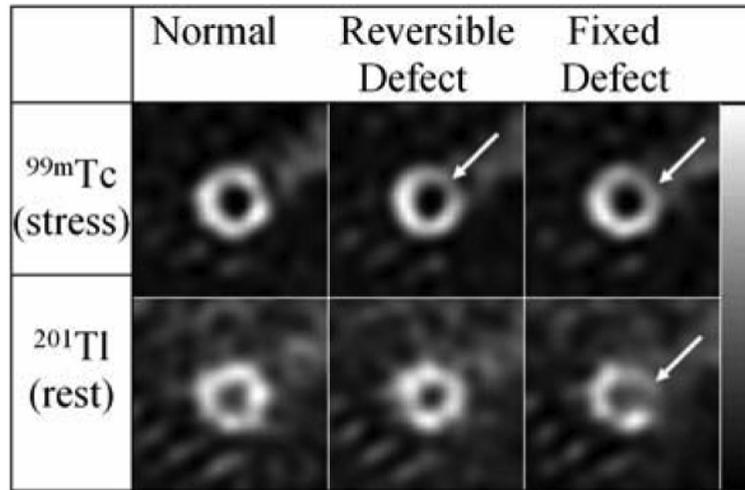
1. Kiat H, Germano G, Friedman J, Vantrain K, Silagan G, Wang FP, Maddahi J, Berman D. Comparative feasibility of separate or simultaneous rest Tl-201 stress technetium-99m-Sestamibi dual-isotope myocardial perfusion SPECT. *J Nucl Med.* 1994; 35:542–548. [PubMed: 8151372]
2. Lowe VJ, Greer KL, Hanson MW, Jaszczak RJ, Coleman RE. Cardiac phantom evaluation of simultaneously acquired dual-isotope rest Tl-201 stress Tc-99m SPECT images. *J Nucl Med.* 1993; 34:1998–2006. [PubMed: 8229250]
3. Moore SC, Syravanh C, Tow DE. Simultaneous SPECT imaging of Tl-201 and Tc-99m using 4 energy windows. *J Nucl Med.* 1993; 34:P188–P188.
4. Nakamura M, Takeda K, Ichihara T, Motomura N, Shimizu H, Saito Y, Nomura Y, Isaka N, Konishi T, Nakano T. Feasibility of simultaneous stress Tc-99m-sestamibi/rest (201)Tl dual-isotope myocardial perfusion SPECT in the detection of coronary artery disease. *J Nucl Med.* 1999; 40:895–903. [PubMed: 10452303]
5. Swets JA, Birdsall TG. The human use of information.3. decision-making in signal-detection and recognition situations involving multiple alternatives. *IRE Trans Inf Theory.* 1956; 2:138–165.
6. Swets, JA.; Pickett, RM. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory.* New York: Academic; 1982.
7. He X, Metz CE, Tsui BMW, Links JM, Frey EC. Three-class ROC analysis—A decision theoretic approach under the ideal observer framework. *IEEE Trans Med Imag.* May; 2006 25(5):571–581.
8. He X, Frey EC. Three-class ROC analysis—The equal error utility assumption and the optimality of three-class ROC surface using the ideal observer. *IEEE Trans Med Imag.* Aug; 2006 25(8):979–986.
9. He X, Frey EC. An optimal three-class linear observer derived from decision theory. *IEEE Trans Med Imag.* Jan; 2007 26(1):77–83.
10. He X, Frey EC. The meaning of the volume under the three-class ROC surface (VUS). *IEEE Trans Med Imag.* May; 2008 27(5):577–588.
11. Van Trees, HL. *Detection, Estimation and Modulation Theory.* New York: Wiley; 1968. p. I
12. Metz, CE. *The optimal decision variable: Likelihood ratio.* Univ. Chicago; Chicago, IL: 2000.
13. Barrett, HH.; Myers, KJ. *Foundations of Image Science.* New York: Wiley; 2003.
14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982; 143:29–36. [PubMed: 7063747]
15. Egan, JP. *Signal Detection Theory and ROC Analysis.* New York: Academic; 1975.
16. Green, DM.; Swets, JA. *Signal Detection Theory and Psychophysics.* Huntington, New York: Krieger; 1973.
17. Lusted LB. Signal detectability and medical decision-making. *Science.* 1971; 171:1217–1217. [PubMed: 5545199]
18. McNeil BJ, Keeler E, Adelstein SJ. Primer on certain elements of medical decision-making. *New Eng J Med.* 1975; 293:211–215. [PubMed: 806804]
19. Swets, JA.; Pickett, RM. *Evaluation of Diagnostic System: Methods from Signal Detection Theory.* New York: Academic; 1982.
20. Chan HP, Sahiner B, Hadjiiski LM, Petrick N, Zhou C. Design of three-class classifiers in computer-aided diagnosis: Monte Carlo simulation study. *Proc SPIE.* 2003; 5032:567–578.
21. Dreiseitl S, Ohno-Machado L, Binder M. Comparing three-class diagnostic tests by three-way ROC analysis. *Med Decision Making.* 2000; 20:323–31.

22. Edwards, DC. Ideal observer estimation and generalized ROC analysis for computer-aided diagnosis. Univ. Chicago; Chicago, IL: 2003.
23. Edwards DC, Lan L, Metz CE, Giger ML, Nishikawa RM. Estimating three-class ideal observer decision variables for computerized detection and classification of mammographic mass lesions. *Med Phys*. 2004; 31:81–90. [PubMed: 14761024]
24. Edwards DC, Metz CE. Restrictions on the three-class ideal observer's decision boundary lines. *IEEE Trans Med Imag*. Dec; 2005 24(12):1566–1573.
25. Edwards DC, Metz CE. Analysis of proposed three-class classification decision rules in terms of the ideal observer decision rule. *J Math Psychol*. 2006; 50:478–487.
26. Edwards DC, Metz CE, Kupinski MA. Ideal observers and optimal ROC hypersurfaces in N-class classification. *IEEE Trans Med Imag*. Jul; 2004 23(7):891–895.
27. Edwards DC, Metz CE, Nishikawa RM. The hypervolume under the ROC hypersurface of “near-guessing” and “near-perfect” observers in N-class classification tasks. *IEEE Trans Med Imag*. Mar; 2005 24(3):293–299.
28. Everson RM, Fieldsend JE. Multi-class ROC analysis from a multi-objective optimisation perspective. *Pattern Recognit Lett*. 2006; 27:918–927.
29. Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach Learn*. 2001; 45:171–186.
30. Heckerling PS. Parametric three-way receiver operating characteristic surface analysis using mathematica. *Med Decision Making*. 2001; 21:409–417.
31. Johnson AJ, Noga AJ, Kosoy O, Lanciotti RS, Johnson AA, Biggerstaff BJ. Duplex microsphere-based immunoassay for detection of anti-West Nile virus and anti-St. Louis encephalitis virus immunoglobulin M antibodies. *Clin iagnostic Lab Immunol*. 2005; 12:566–574.
32. Kijewski MF, Swensson RG, Judy PF. Analysis of rating data from multiple-alternative tasks. *J Math Psychol*. 1989; 33:428–451.
33. Mossman D. Three-way ROCs. *Med Decision Making*. 1999; 19:78–89.
34. Nakas CT, Yiannoutsos CT. Ordered multiple-class ROC analysis with continuous measurements. *Stat Med*. 2004; 23:3437–3449. [PubMed: 15505886]
35. Patel, AC.; Markey, MK. *SPIE Med Imag*. Vol. 5749. San Diego, CA: 2005. Comparison of three-class classification performance metrics: A case study in breast cancer CAD; p. 581–589.
36. Sahiner B, Chan HP, Hadjiiski LM. Performance analysis of three-class classifiers: Properties of a 3-D ROC surface and the normalized volume under the surface for the ideal observer. *IEEE Trans Med Imag*. Feb; 2008 27(2):215–227.
37. Sahiner B, Chan HP, Hadjiiski LM. Performance analysis of 3-class classifiers: Properties of the 3-D ROC surface and the normalized volume under the surface. *Proc SPIE*. 2006; 6146:567–578.
38. Scurfield BK. Multiple-event forced-choice tasks in the theory of signal detectability. *J Math Psychol*. 1996; 40:253–269. [PubMed: 8979976]
39. Scurfield BK. Generalization of the theory of signal detectability to n-event m-dimensional forced-choice tasks. *J Math Psychol*. 1998; 42:5–31. [PubMed: 9606159]
40. Xiong CJ, van Belle G, Miller JP, Morris JC. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Stat Med*. 2006; 25:1251–1273. [PubMed: 16345029]
41. Yeung KY, Bumgarner RE, Raftery AE. Bayesian model averaging: Development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics*. 2005; 21:2394–2402. [PubMed: 15713736]
42. He X, Frey EC. The meaning of the volume under the three-class ROC surface (VUS). *IEEE Trans Med Imag*. May; 2008 27(5):577–588.
43. He X, Frey E, Segars W, Gilland K, Tsui B. An improved phantom population for evaluation of myocardial perfusion SPECT imaging. *J Nucl Med*. 2002; 43:9–10.
44. He X, Frey EC, Links JM, Gilland KL, Segars WP, Tsui BMW. A mathematical observer study for the evaluation and optimization of compensation methods for myocardial SPECT using a phantom population that realistically models patient variability. *IEEE Trans Nucl Sci*. Feb; 2004 51(1):218–224.

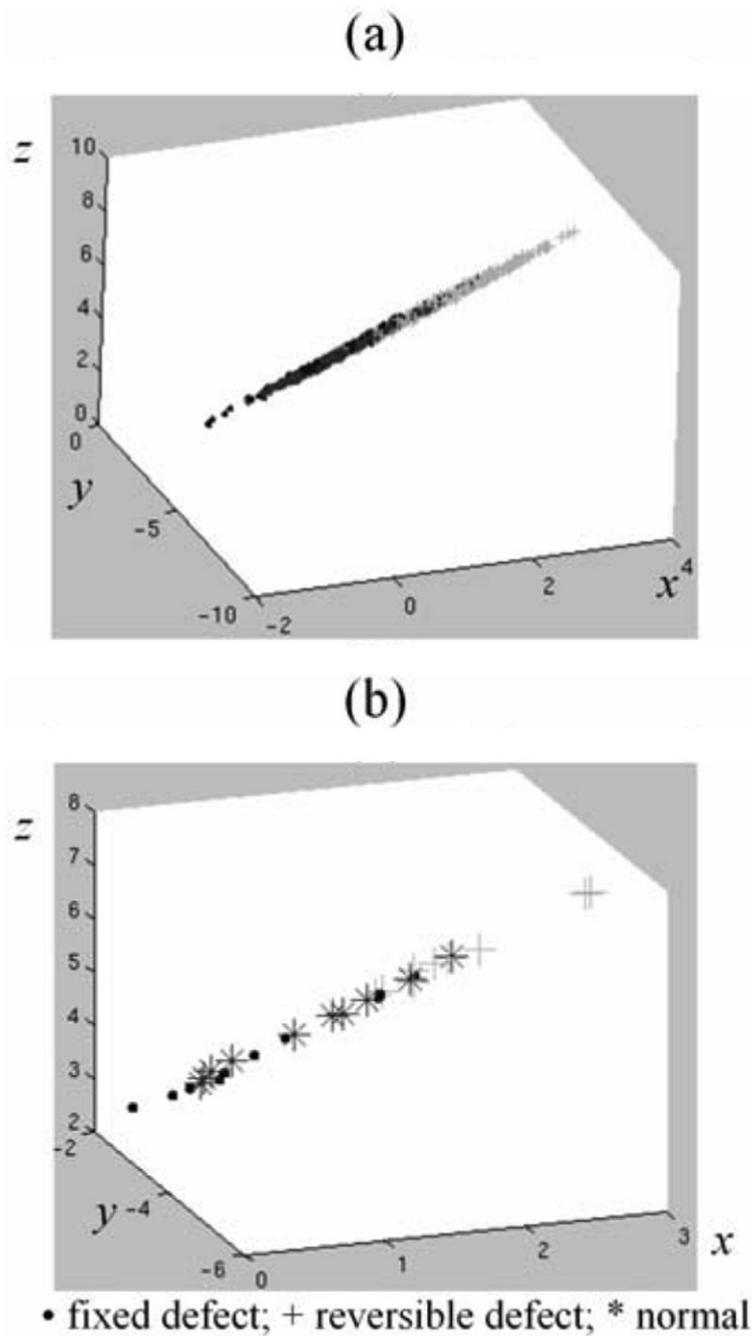
45. Segars WP, Tsui BM, Lalush DS, Frey EC, King MA, Manocha D. Development and application of the new dynamic nurbs-based cardiac-torso (NCAT) phantom. *J Nucl Med.* 2001; 42:7.
46. Barclay AB, Eisner RL, DiBella EV. PET thorax model database [Online]. Available: <http://www.emory.edu/CRL/abb/thorax-model/Contents.html>
47. He X, Links JM, Gilland KL, Tsui BMW, Frey EC. Comparison of 180 degrees and 360 degrees acquisition for myocardial perfusion SPECT with compensation for attenuation, detector response, and scatter: Monte carlo and mathematical observer results. *J Nucl Cardiol.* 2006; 13:345–353. [PubMed: 16750779]
48. Song X, Frey EC, Segars WP, Tsui BMW. A mathematical observer study for evaluation of a model-based compensation method for crosstalk in simultaneous dual isotope SPECT. *IEEE Med Imag Conf.* 2003; 4:2521–2525.
49. SimSET Version 2.6 Univ. Washington, 2004 [Online]. Available: [http://depts.Washington.edu/simsimset/html/simset\\_home.html](http://depts.Washington.edu/simsimset/html/simset_home.html)
50. Song X, Segars WP, Du Y, Tsui BMW, Frey EC. Fast modelling of the collimator-detector response in monte Carlo simulation of SPECT imaging using the angular response function. *Phys Med Biol.* 2005; 50:1791–1804. [PubMed: 15815096]
51. Frey EC, Gilland KL, Tsui BM. Application of task-based measures of image quality to optimization and evaluation of three-dimensional reconstruction-based compensation methods in myocardial perfusion SPECT. *IEEE Trans Med Imag.* Sep; 2002 21(9):1040–1050.
52. Sankaran S, Frey EC, Gilland KL, Tsui BMW. Optimum compensation method and filter cutoff frequency in myocardial SPECT: A human observer study. *J Nucl Med.* 2002; 43:432–438. [PubMed: 11884505]



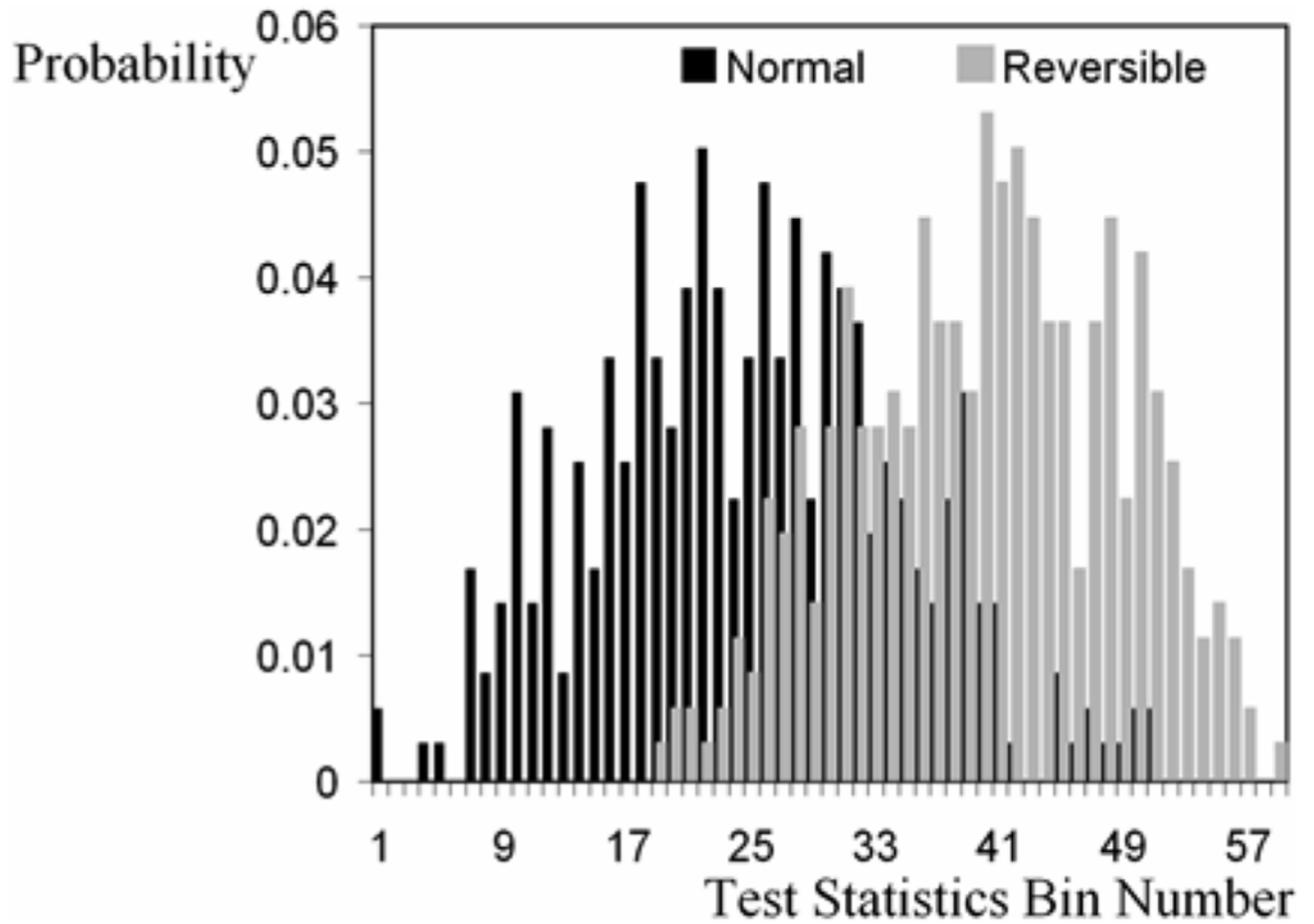
**Fig. 1.** (a) Decision space of the three-class ROC analysis. Differently shaded ellipses represent the decision variable distributions of the three classes, (b) Example of the three-class ROC surface.



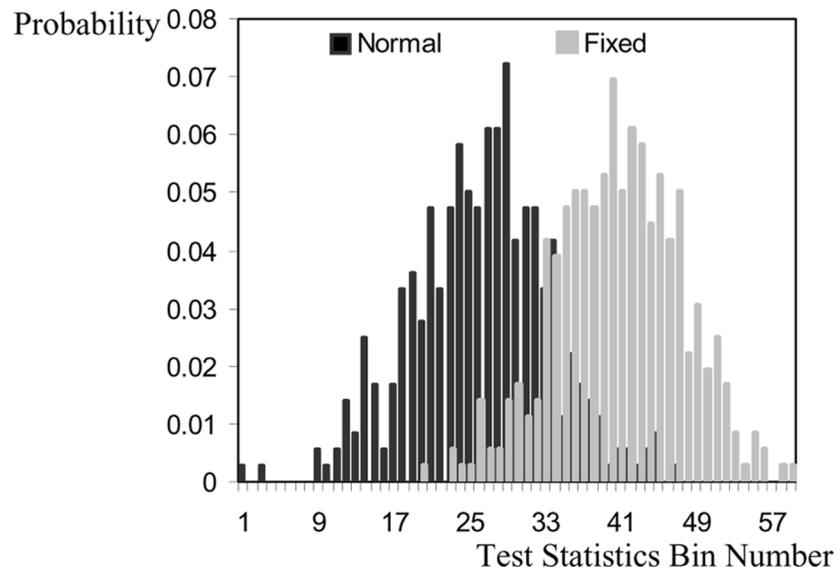
**Fig. 2.**  
Example of a triplet of dual-isotope MPS images (white arrow indicates the location of the defect).



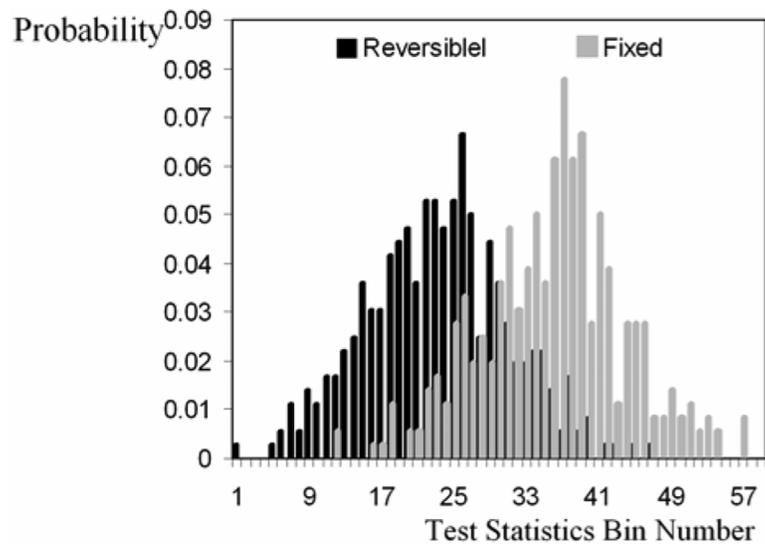
**Fig. 3.** Result of three-template test for the data treatment in Table II. At one orientation, all points align in a line, suggesting a plane in the 3-D space, (a)  $(x, y, z) = (\vec{w}_{23}^t \vec{v}, \vec{w}_{13}^t \vec{v}, \vec{w}_{21}^t \vec{v})$  triplets from all 360 triplets of feature vectors, (b)  $(\vec{w}_{23}^t \vec{v}, \vec{w}_{13}^t \vec{v}, \vec{w}_{21}^t \vec{v})$  triplets from only 10 triplets of feature vectors for a clearer view.



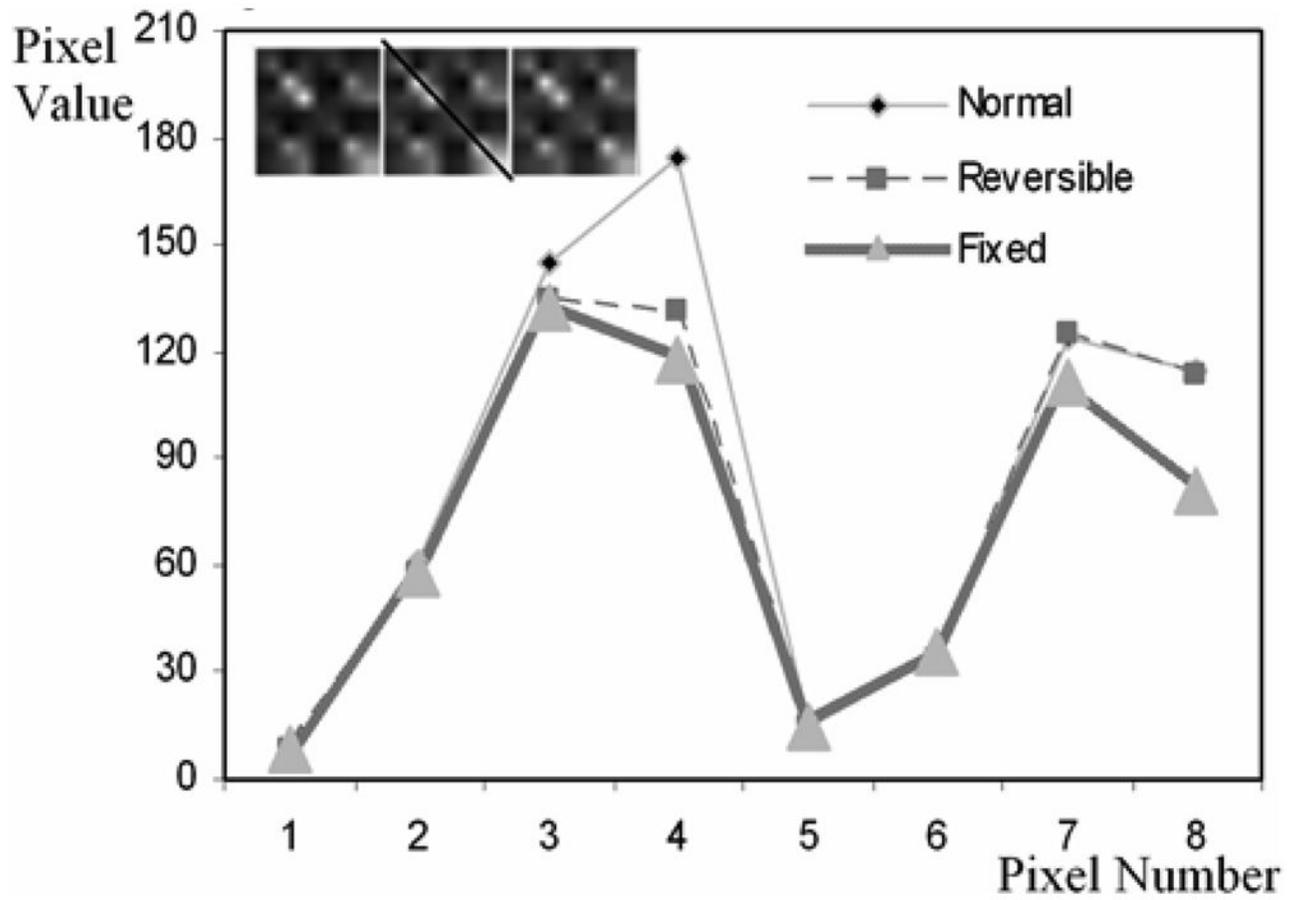
**Fig. 4.** Histograms of the binary test statistics between Normal class and Reversible class.



**Fig. 5.**  
Histograms of the binary test statistics between Normal class and Fixed class.



**Fig. 6.** Histograms of the binary test statistics between Reversible class and fixed class.



**Fig. 7.**  
Profiles of the diagonal of the covariance matrixes.

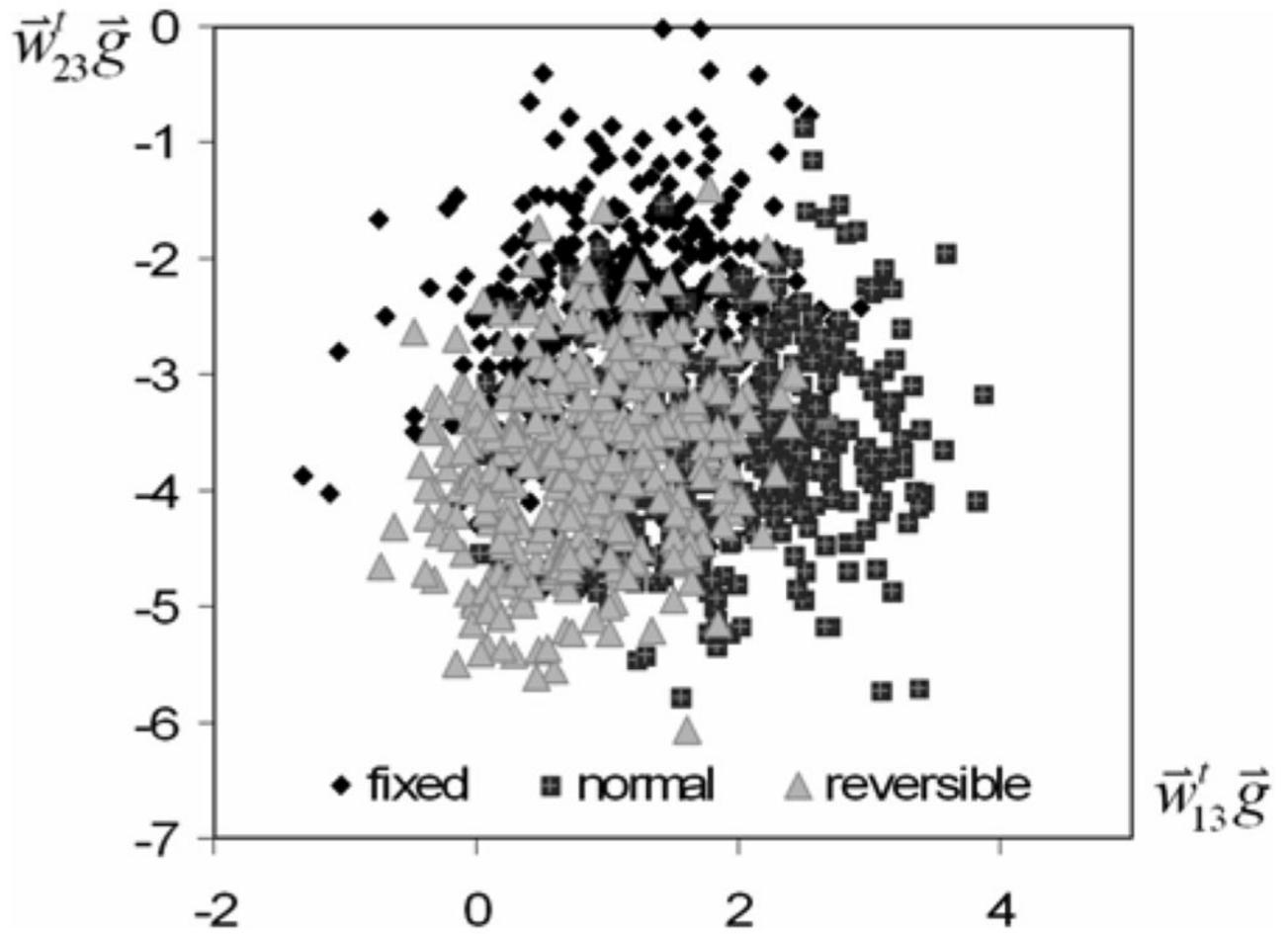
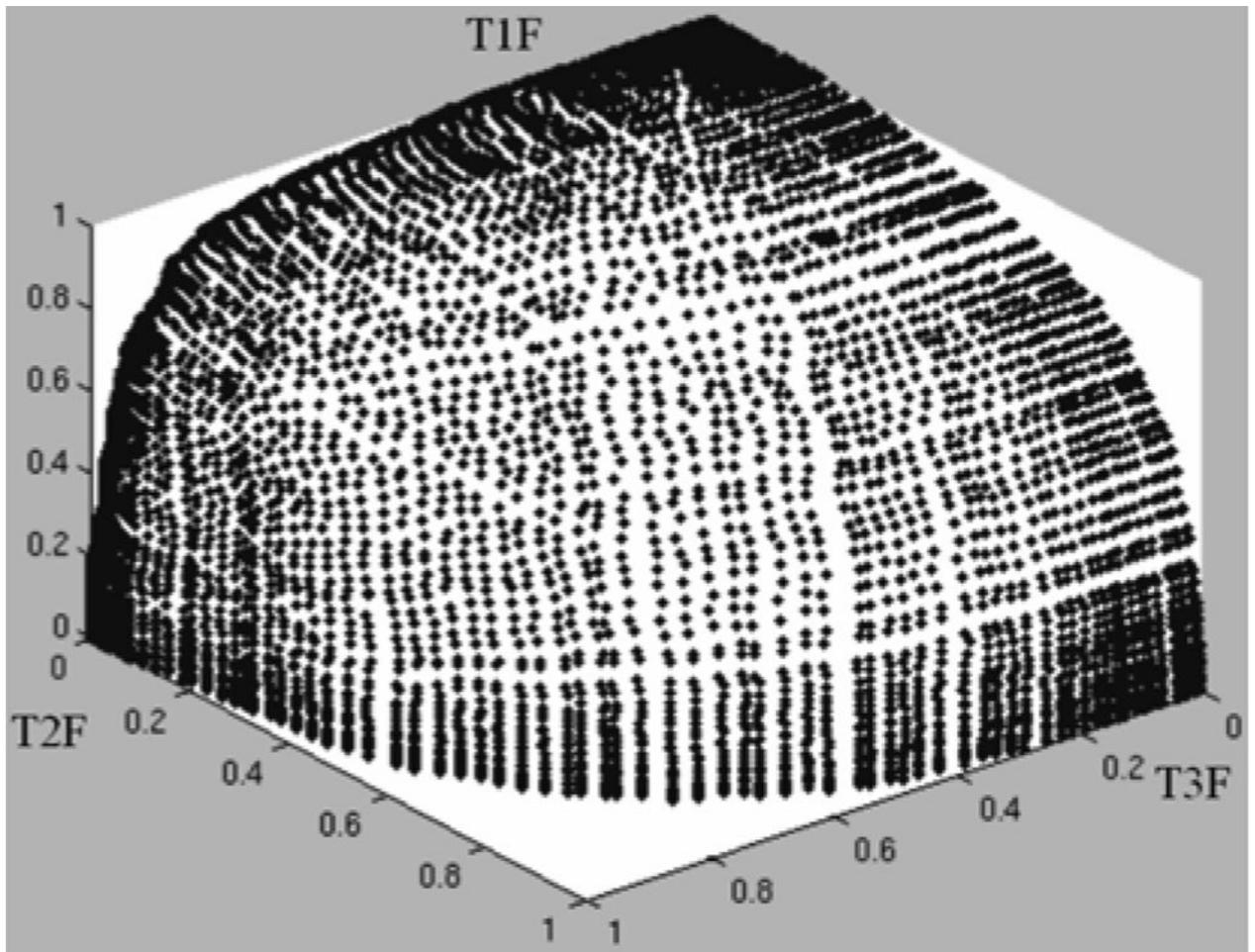
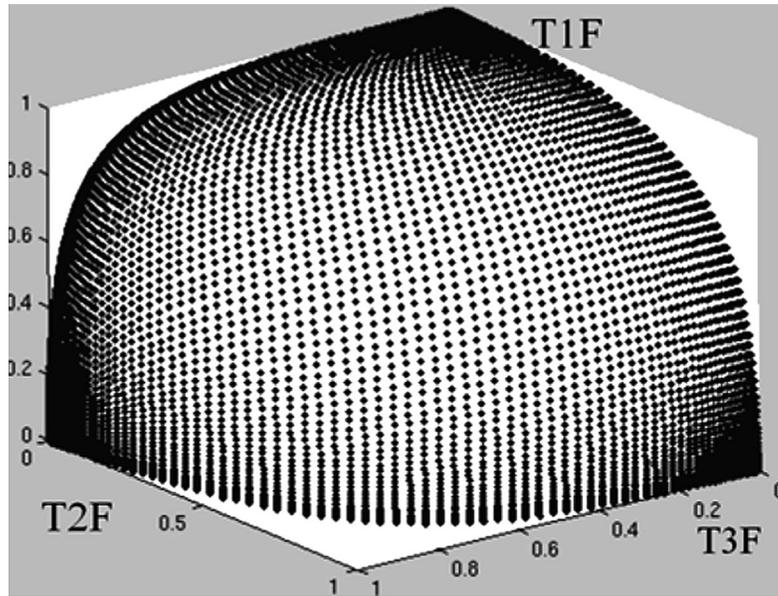


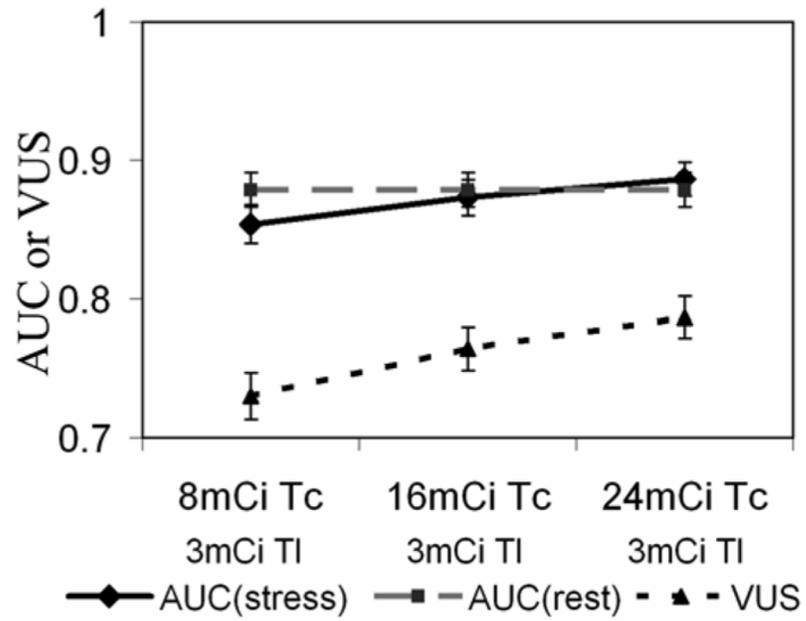
Fig. 8.  
Decision plane obtained for simulated dual-isotope MPS images.



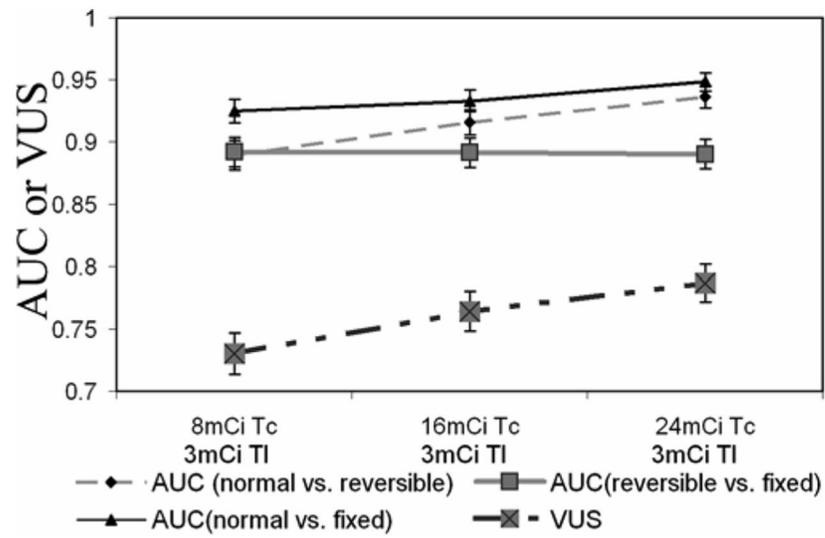
**Fig. 9.**  
ROC surface obtained for simulated dual-isotope MPS images, where  $VUS = 0.703$ .



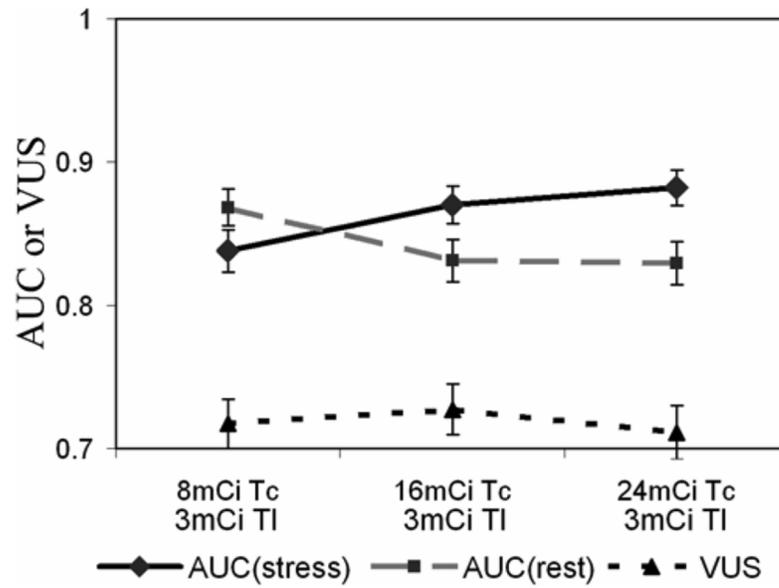
**Fig. 10.** ROC fitted surface obtained for simulated dual-isotope MPS images, where  $VUS = 0.704$ .



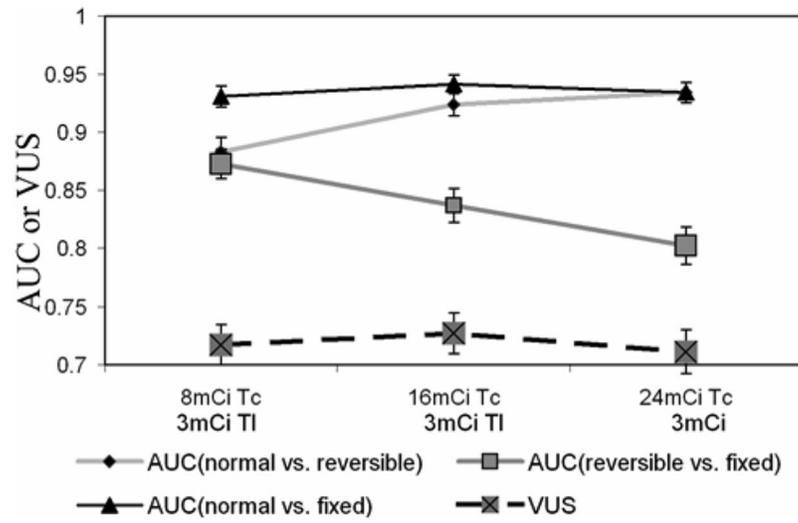
**Fig. 11.** Comparison of binary observer performance of stress ( $^{99m}\text{Tc}$ ) and rest ( $^{201}\text{Tl}$ ) images with three-class observer performance for separate acquisition.



**Fig. 12.** Comparison of pairwise binary observer performance with three-class observer performance for separate acquisition.



**Fig. 13.** Comparison of binary observer performance of stress ( $^{99m}\text{Tc}$ ) and rest ( $^{201}\text{Tl}$ ) images with three-class observer performance for separate acquisition.



**Fig. 14.** Comparison of pairwise binary observer performance with three-class observer performance for simultaneous acquisition.

**TABLE I**

Three Classes in the Simultaneous Dual-Isotope MPS imaging

	<b>Class 1</b>	<b>Class 3</b>	<b>Class 2</b>
	<b>Normal</b>	<b>Reversible defect</b>	<b>Fixed defect</b>
$^{99m}\text{Tc}$ (stress)	No *	Yes **	Yes
$^{201}\text{Tl}$ (rest)	No	No	Yes

\* No denotes for defect absence

\*\* Yes denotes for defect presence

**TABLE II**

Data Treatment Used in This 3-CHO Study

<b>Acquisition strategy: Simultaneous (with crosstalk contamination)</b>				
	<b>Injected activity</b>	<b>Reconstruction method</b>	<b>Iteration</b>	<b>Cutoff frequency</b>
Stress ( $^{99m}\text{Tc}$ )	8mCi	OSEM-ADS	13	0.22/pix
Rest ( $^{201}\text{Tl}$ )	3mCi	OSEM-ADS	11	0.30/pix

TABLE III

Results of the Plane Fitting for the Three-Template Test

	b	c	d	RSMRE	L
Ideal	-1	-1	0	0	
Class 1	-0.9285	-0.9778	0.0072	0.1282	360
Class 2	-0.9201	-0.9832	0.0685	0.1815	360
Class 3	-0.9171	-0.9913	0.1151	0.1651	360
All	-0.9269	-0.9842	0.0410	0.2769	360×3

**TABLE IV**

Normality Test of the Test Statistics in the Binary ROC Analysis Involving Normal Class and Reversible Class

	<b>H* (P=0.01)</b>	<b>P</b>
Normal	0	0.188823
Reversible	0	0.023157

\* H=0 Do not reject the null hypothesis at significance level P=0.01

\* H=1 Reject the null hypothesis at significance level P=0.01

**TABLE V**

Normality Test of the Test Statistics in the Binary ROC Analysis Involving Normal Class and Fixed Class

	<b>H*</b> (P=0.01)	<b>P</b>
Normal	0	0.328528
Fixed	0	0.902211

\* Same as in Table 4.

**TABLE VI**

Normality Test of the Test Statistics in the Binary ROC Analysis Involving Reversible Class and Fixed Class

	<b>H*</b> ( <b>P=0.01</b> )	<b>P</b>
Reversible	0	0.798496
Fixed	0	0.555546

\* Same as in Table 4.

TABLE VII

Data Treatment Used in This Experiment

<b>Acquisition strategy:</b>				
<b>Simultaneous (with crosstalk contamination)</b>				
<b>Separate acquisition</b>				
	<b>Injected Activity Combination</b>	<b>Reconstruction method</b>	<b>Iterations</b>	<b>Cutoff frequency</b>
Stress ( $^{99m}\text{Tc}$ )	8mCi	OSEM-ADS	10	0.20/pix
Rest ( $^{201}\text{Tl}$ )	3mCi	OSEM-ADS	10	0.22/pix
Stress ( $^{99m}\text{Tc}$ )	16mCi	OSEM-ADS	10	0.20/pix
Rest ( $^{201}\text{Tl}$ )	3mCi	OSEM-ADS	10	0.22/pix
Stress ( $^{99m}\text{Tc}$ )	24mCi	OSEM-ADS	10	0.20/pix
Rest ( $^{201}\text{Tl}$ )	3mCi	OSEM-ADS	10	0.22/pix