Published in final edited form as:
IEEE Trans Med Imaging. 2010 January ; 29(1): 206–215. doi:10.1109/TMI.2009.2034516.

Three-Class ROC Analysis—Toward a General Decision Theoretic Solution

Xin He,
Department of Radiology, Johns Hopkins School of Medicine, Baltimore, MD 21287 USA (xinhe@jhmi.edu)

Brandon D. Gallas, and
DIAM/OSEL/CDRH, Food and Drug Administration, Silver Spring, MD, 20993 USA (brandon.gallas@fda.hhs.gov)

Eric C. Frey
Department of Radiology, Johns Hopkins School of Medicine, Baltimore, MD 21287 USA (efrey@jhmi.edu)

Abstract

Multiclass receiver operating characteristic (ROC) analysis has remained an open theoretical problem since the introduction of binary ROC analysis in the 1950s. Previously, we have developed a paradigm for three-class ROC analysis that extends and unifies decision theoretic, linear discriminant analysis, and probabilistic foundations of binary ROC analysis in a three-class paradigm. One critical element in this paradigm is the equal error utility (EEU) assumption. This assumption allows us to reduce the intrinsic space of the three-class ROC analysis (5-D hypersurface in 6-D hyperspace) to a 2-D surface in the 3-D space of true positive fractions (sensitivity space). In this work, we show that this 2-D ROC surface fully and uniquely provides a complete descriptor for the optimal performance of a system for a three-class classification task, i.e., the triplet of likelihood ratio distributions, assuming such a triplet exists. To be specific, we consider two classifiers that utilize likelihood ratios, and we assumed each classifier has a continuous and differentiable 2-D sensitivity-space ROC surface. Under these conditions, we proved that the classifiers have the same triplet of likelihood ratio distributions if and only if they have the same 2-D sensitivity-space ROC surfaces. As a result, the 2-D sensitivity surface contains complete information on the optimal three-class task performance for the corresponding likelihood ratio classifier.

Keywords

Extended receiver operating characteristic (ROC) analysis; ROC analysis; three-class ROC analysis

I. Introduction

Many medical diagnostic problems involve more than two diagnostic alternatives. For example, in mammography, patients can be classified as having no tumors or benign or malignant tumors. Similarly, in myocardial perfusion imaging, patients can have normal perfusion or fixed or reversible perfusion defects. To effectively evaluate and optimize diagnostic techniques for such multialternative tasks, a general multiclass receiver operating

Correspondence to: Xin He.

characteristic (ROC) analysis method is required, but has remained an open problem ever since the introduction of binary ROC analysis in 1950s [1], [2].

According to the current understanding of three-class classification performance, task performance can only be described with full generality by a 5-D hypersurface in a 6-D space of decision outcome fractions, which are most often taken to be the six false decision outcome fractions [3], [4]. Previously, we have proposed a solution to the three-class ROC analysis problem that extends and unifies the decision theoretic, linear discriminant analysis (LDA) and probabilistic foundations of binary ROC analysis in a three-class paradigm [2]-[5]. This method was originally derived under a decision theoretic framework by assuming that the utilities for incorrect classifications were the same under a given class. This assumption was termed the equal error utility assumption (EEU). Under this assumption, we demonstrated that the task performance can be assessed in terms of a 2-D surface in the 3-D space of sensitivities (true positive fractions).

The goal of this work is to build a foundation for investigating whether the EEU assumption yields a limited three-class ROC analysis or is truly utility independent under decision theory. In the following, we first review previous work on decision-theory based three-class ROC analysis to provide a foundation for this work. Next, we present a particular perspective for interpreting three-class task performance. To be specific, we argue that optimal three-class task performance can be completely characterized by the triplet of likelihood distributions, and that this triplet is a full descriptor of general three-class classification performance. With this foundation, we prove that a 2-D ROC surface generated from the likelihood ratio distributions uniquely determines that triplet of likelihood ratio distributions, making such a 2-D ROC surface a full-descriptor of optimal three-class classification performance.

II. Background

ROC analysis has been widely accepted as a way to describe binary classification performance, and binary ROC analysis methodology is well established. To evaluate three-class classification performance, three-class ROC analysis methodology is needed. Many three-class ROC analysis methods have been proposed by extending binary ROC analysis in various ways [1]-[29]. Among these, the most appealing approaches have been the decision theoretic ones due to their strong foundation in decision theory [12], [19]-[27]. In the following, we briefly introduce the fundamentals of decision theoretic three-class ROC analysis.

A. General Three-Class ROC Analysis

This approach originated with Metz's reformulation [12] to a three-class paradigm of the decision theoretic framework proposed by Van Trees [30]. Metz's general three-class ROC theory rigorously formulates the basic decision theoretic aspects of three-class ROC analysis. In a three-class classification task, whose decision table is shown in Table I, there are nine diagnostic accuracy fractions and three relationships between them, namely

$$T1F+F21F+F31F=1 \quad (1)$$

$$F12F+T2F+F32F=1 \quad (2)$$

and

$$F13F+F23F+T3F=1 \quad (3)$$

where *TiF* is the probability that class *i* is correctly classified, and *FijF* is the probability that a Class *i* decision is made while the underlying truth is class *j*. Thus, of the nine diagnostic accuracy fractions, six are needed to fully describe three-class classification performance. Therefore, we see that when moving from 2 to 3 classes, the total number of fractions needed to fully describe the classification performance increases from 2 to 6. Thus, the binary ROC curve is a 1-D curve in a 2-D ROC space and three-class ROC hypersurface is proposed to be a 5-D hypersurface in a 6-D space.

In the following, the set of {*TiF*} will be referred to as the set of sensitivities and the set of {*FijF*} will be referred to as the set of false decision outcome fractions. Under Metz's three-class decision theoretic formulation, it has been proved that optimal decision variables for three-class classification are two likelihood ratios [12], [20], i.e.,

$$\Lambda_{13}=\Lambda_{13}\left(\vec{g}\right)=\frac{f_1\left(\vec{g}\right)}{f_3\left(\vec{g}\right)} \quad (4)$$

and

$$\Lambda_{23}=\Lambda_{23}\left(\vec{g}\right)=\frac{f_2\left(\vec{g}\right)}{f_3\left(\vec{g}\right)} \quad (5)$$

where $\vec{g}$ is the data vector, $f_i(\vec{g})$ is the likelihood function of the *i*th class, and Λ_{i3} ($i = 1, 2$) is the likelihood ratio of Class *i* to Class 3.

The decision space is thus spanned by two likelihood ratios. Given a set of prior information (in the form of class prevalence and utilities of each of the nine possible decision outcomes), a set of decision rules that maximize the expected utility is represented by three rays originating from a common point, known as the critical point. These rays partition the decision space into three regions. Each of these regions corresponds to a region where one of the decisions is made. The rays delineating these regions are expressed as [12]-[20],

$$[U_{11}-U_{21}]P_1\Lambda_{13}+[U_{13}-U_{23}]P_3=[U_{22}-U_{12}]P_2\Lambda_{23} \quad (6)$$

$$[U_{11}-U_{31}]P_1\Lambda_{13}+[U_{13}-U_{33}]P_3=[U_{32}-U_{12}]P_2\Lambda_{23} \quad (7)$$

and

$$[U_{21}-U_{31}]P_1\Lambda_{13}+[U_{23}-U_{33}]P_3=[U_{32}-U_{22}]P_2\Lambda_{23} \quad (8)$$

where U_{ij} is the utility for deciding Class *i* when the truth is Class *j*, and P_i is the prior prevalence of the of the *i*th class. An example of the likelihood ratio decision plane is shown in Fig. 1.

The three lines are determined by six parameters, and since the three lines intersect at one point, there are a total of 5 parameters required. As a result, to fully describe three-class task performance, Metz and Edwards proposed a 5-D ROC hypersurface in a 6-D hyperspace spanned by the six false decision outcome fractions. This 5-D hypersurface in the 6-D hyperspace is termed the general three-class ROC analysis method [12], [19]-[27].

B. Practical Three-Class ROC Analysis

We have previously proposed a decision theoretic three-class ROC analysis method. In the derivation of this method, we reduced the dimensionality of the three-class problem by making the equal-error utility (EEU) assumption [31]. This assumption states

$$U_{31}=U_{21}, U_{32}=U_{12}, \text{ and } U_{13}=U_{23}. \quad (9)$$

As a result, the number of degrees-of-freedom of the decision structure is reduced to 2. The resulting decision spaces and structures spanned by the likelihood ratio and log likelihood ratios are shown in Fig. 2. In the log likelihood ratio decision space (right plot), the shape of the decision structure does not vary as it moves across the decision space due to the EEU assumption. To be specific, the angles of the three lines with respect to the *x*-axis are always 0°, 90°, and 45°. In the likelihood decision space (left plot), the boundary between class 1 and 2 varies, but the other two are fixed as vertical and horizontal. Moving the decision structure across the decision space and computing the set of sensitivities, ($T1F$, $T2F$, $T3F$), traces out a 2-D three-class ROC surface in the 3-D space spanned by $\{TiF\}$. An example of the resulting three-class ROC surface is shown in Fig. 3.

The volume under the three-class ROC surface (VUS) was proposed as a figure-of-merit (FOM) for task performance under the EEU [31]. The VUS is analogous to the area under the ROC curve (AUC) in the two-class case: it is a FOM for EEU task performance in the case where the surfaces or curves do not cross.

We have further investigated this 2-D three-class ROC surface and demonstrated that it contains all the optimal operating points that maximize the probabilities of making correct decisions for all possible combinations of prior prevalence (i.e., disease prevalence); it contains all the optimal operating points that satisfy Neyman–Pearson criterion in a sense of maximizing one sensitivity given the other two; and it contains the optimal operating point that satisfies the maximum likelihood criterion [32]. In addition, when the data follow Gaussian distributions with equal covariances, the signal-to-noise ratios (SNRs) of the test statistics between each pair of the classes are simultaneously maximized [33]. Finally, we have shown that the VUS value is equivalent to the percent correct in a three-class categorization procedure—a relationship that provides a foundation for nonparametric statistical analysis methods for three-class ROC analysis as well as a practical method for data collection in reader studies [34].

III. Task Performance, Decision Variables, and Decision Rules

In this section, we present a particular perspective for viewing three-class classification performance and its relationship to the decision variables and decision rules. We then discuss the necessary properties of a general task performance display or FOM.

A. Pair/Triplet of Likelihood Ratio Distributions as a Complete Descriptor of Binary/Three-Class Optimal Task Performance

In a classification task, the decision space is a space spanned by decision variables and partitioned according to decision rules. Fig. 4 schematically shows examples of a pair of likelihood ratio distributions in a binary decision space and a triplet of likelihood distributions in a three-class decision space. Shaded ellipses schematically represent distributions of the likelihood ratios of the three classes. Given the triplet of likelihood ratio distributions, the optimal task performance is known exactly. That is, for any set of prior information (prevalences and utilities), we can compute the optimal decision structure and use it to evaluate the nine decision outcome fractions.

While the triplet of likelihood ratio distributions is a complete descriptor of task performance, it is not practically useful: it is very difficult to compare two systems directly via their likelihood ratio distributions on the likelihood ratio decision space. As a result, we believe the goal of ROC analysis is to provide a standard for summarizing and comparing the likelihood ratio distributions of different systems.

Since the triplet of likelihood ratio distributions provides a complete descriptor of three-class task performance, a necessary condition for an ROC surface to be a general descriptor is that there is a one-to-one relationship between the ROC surface and the triplet of likelihood ratio distributions. In other words, if each unique ROC surface is uniquely related to a unique set of likelihood ratio distributions, then the ROC surface can be said to be general in a decision theoretic sense.

Some readers might wonder whether binary ROC analysis fits the above conditions, since it is well known that different pairs of rating distributions would yield the same ROC curve. In such a case, the decision variables are monotonic transformations of each other. However, only one transformation will yield the unique pair of corresponding likelihood ratio distributions due to the properties of likelihood ratios, to be described in Section IV. It is this constraint, relating an ROC surface to its equivalent ideal observer, that we shall use to uniquely relate an ROC surface to the triplet of likelihood ratio distributions.

B. Role of Decision Rules in Optimal Task Performance Assessment

The decision structure, which represents the decision rules, specifies how to partition the decision space into regions corresponding to the possible decisions. The 5-D hypersurface is obtained using the 5-D decision structure (Fig. 1) to partition the likelihood ratio decision space. It provides one way of describing the triplet of likelihood ratio distributions, and has been considered to be a complete descriptor of three-class task performance [19]-[27]. The 2-D ROC surface is obtained by using the 2-D decision structure to partition the likelihood ratio decision space (Fig. 2). It provides another way of describing the triplet of likelihood ratio distributions and has been considered to be an incomplete descriptor of three-class task performance due to the use of the EEU assumption [20].

The above two examples reveal the roles of decision rules in task performance description. They provide ways to convert the triplet of likelihood ratio distributions—a complete descriptor of task performance—to a hypersurface in a 3-D or 6-D space spanned by different sets of decision outcome fractions, with the goal of finding a convenient way to summarize task performance. Note that both the 2-D and 5-D decision structures use the same triplet of likelihood ratio distributions. As long as this conversion is one-to-one or reversible, there is no loss of information about task performance.

C. Necessary Properties of a General Three-Class Task Performance Descriptor in the Decision Outcome Fraction Space

As discussed above, the goal of ROC analysis is to convert the likelihood ratio distributions to a hypersurface in a decision outcome fraction space with the goal of finding a tractable descriptor of optimal task performance that can be used to compare systems. Let us now consider some of the essential (necessary but insufficient) properties of such a conversion. (Note that we are elucidating these points here, but will not prove them or examine them further.)

1. It must provide a way to determine the conditional probabilities for all possible utility combinations. For example, a binary ROC curve describes the sensitivity and specificity pairs of all possible utility combinations. In the proposed 2-D decision model for three-class ROC analysis, the 2-D ROC surface describes the conditional

probabilities for all possible utility combinations that satisfy the EEU assumption. A general three-class descriptor must also have the ability to describe the conditional probabilities in cases when EEU is not valid.

2. The general descriptor should not be biased toward any class. This means that the system should be independent of the class labels. For example, in evaluating binary classifiers, the ranking of the systems using the ROC curve is not affected by calling positive cases Class 1 or Class 2.
3. A general metric summarized from a general descriptor must rank ideal observer performance as the optimal one, i.e., the ideal observer must have better performance than any other observers.

IV. Theory

We propose the following conjecture.

Conjecture—There is a unique triplet of likelihood ratios that produces a given well-behaved 2-D ROC surface.

In the following, we first introduce the properties of likelihood ratio distributions, which are keys to the proof, and then prove the conjecture by demonstrating the unique relationship between the 2-D ROC surface and triplet of likelihood ratio distributions, assuming such a triplet exist exists.

A. Properties of Likelihood Ratio Distributions

Binary Classification—In the following we review several properties of likelihood ratios [35] that are used in the proofs in this paper. The likelihood ratio between Classes *a* and *b* is defined as

$$\Lambda=\Lambda\left(\vec{g}\right)=\frac{\mathrm{f}_a\left(\vec{g}\right)}{\mathrm{f}_b\left(\vec{g}\right)} \quad (10)$$

where $\vec{g}$ is the data vector, and $\mathrm{f}_a(\vec{g})$ and $\mathrm{f}_b(\vec{g})$ are the likelihood functions under Class *a* and *b* hypotheses, respectively. Likelihood ratio distributions have the property that *the likelihood ratio of the likelihood ratio is the likelihood ratio* [35], i.e.,

$$\Lambda=\Lambda\left(\vec{g}\right)=\frac{g_a(\Lambda)}{g_b(\Lambda)} \quad (11)$$

where $\mathrm{g}_a(\Lambda)$ and $\mathrm{g}_b(\Lambda)$ are the distributions of the likelihood ratios under Class *a* and *b* hypotheses.

Three-Class Classification—In three-class ROC analysis, previous work has shown that the optimal decision variables are the two likelihood ratios as expressed in (4) and (5) [12], [20], [31]. Similarly, we can extend the relationship in (11) to three-class

$$\Lambda_{i3}=\frac{g_i(\Lambda_{13},\Lambda_{23})}{g_3(\Lambda_{13},\Lambda_{23})} \quad (12)$$

where $\mathrm{g}_i()$ ($i = 1, 2$) is the joint distribution of the likelihood ratios under *i*th hypothesis. The proof for (12) can be found in [36].

B. Unique Relationship Between the Triplet of Likelihood Ratio Distributions and the 2-D ROC Surface

Consider two systems with log likelihood ratios given by $(\lambda_{13}, \lambda_{23}) = (\log\Lambda_{13}, \log\Lambda_{23})$ for System 1 and $(\lambda'_{13}, \lambda'_{23}) = (\log\Lambda'_{13}, \log\Lambda'_{23})$ for System 2. Let each system have continuous and differentiable 2-D sensitivity-space ROC surfaces. If the triplets of log likelihood ratio distributions are identical in System 1 and System 2, it is obvious that the two systems have identical 2-D ROC surfaces. Below we prove that if two systems have identical 2-D ROC surfaces, which are continuous and differentiable, the two systems must have identical triplets of likelihood ratio distributions. In other words, the two systems are identical in terms of all the decisions they produce.

Since the 2-D ROC surface is continuous and differentiable everywhere and these two systems have identical 2-D ROC surfaces, the corresponding operating points (sensitivities) on the surfaces must have identical coordinates and identical derivatives. Mathematically, we have the following.

1. For each $(\lambda_{13}, \lambda_{23})$, there is an $(\lambda'_{13}, \lambda'_{23})$ such that $\mathrm{TiF}(\lambda_{13}, \lambda_{23}) = \mathrm{TiF}'(\lambda'_{13}, \lambda'_{23})$, where $i = 1, 2, 3$.
2. For each operating point on the 2-D ROC surface, i.e., $\{T1F, T2F, T3F\}$ obtained from the former system and $\{T1F', T2F', T3F'\}$ obtained from the latter, the derivatives must be the same. Thus, for $(\lambda_{13}, \lambda_{23})$ and $(\lambda'_{13}, \lambda'_{23})$ above we have

$$\left.\frac{\partial T3F}{\partial T2F}\right|_{\{\mathrm{TiF}(\lambda_{13},\lambda_{23})\}} = \left.\frac{\partial T3F'}{\partial T2F'}\right|_{\{\mathrm{TiF}'(\lambda'_{13},\lambda'_{23})\}}$$

$$\left.\frac{\partial T3F}{\partial T1F}\right|_{\{\mathrm{TiF}(\lambda_{13},\lambda_{23})\}} = \left.\frac{\partial T3F'}{\partial T1F'}\right|_{\{\mathrm{TiF}'(\lambda'_{13},\lambda'_{23})\}}$$

and

$$\left.\frac{\partial T2F}{\partial T1F}\right|_{\{\mathrm{TiF}(\lambda_{13},\lambda_{23})\}} = \left.\frac{\partial T2F'}{\partial T1F'}\right|_{\{\mathrm{TiF}'(\lambda'_{13},\lambda'_{23}).\}}$$

In Appendix II, we provide a mathematical derivation for computing $\partial T2F/\partial T1F$, $\partial T3F/\partial T1F$ and $\partial T3F/\partial T2F$. Following Appendix II, the partial derivatives of the first system are

$$\left.\frac{\partial T3F}{\partial T1F}\right|_{\{\mathrm{TiF}(\lambda_{13},\lambda_{23})\}} = -e^{-\lambda_{13}}$$

and

$$\left.\frac{\partial T3F}{\partial T2F}\right|_{\{\mathrm{TiF}(\lambda_{13},\lambda_{23})\}} = -e^{-\lambda_{23}}. \quad (13)$$

Similarly, for the second system we have

$$\left.\frac{\partial T3F'}{\partial T1F'}\right|_{\{\mathrm{TiF}'(\lambda'_{13},\lambda'_{23})\}} = -e^{-\lambda'_{13}}$$

and

$$\left.\frac{\partial T3F'}{\partial T2F'}\right|_{\{\mathrm{TiF}'(\lambda'_{13},\lambda'_{23})\}} = -e^{-\lambda'_{23}}. \quad (14)$$

Combining (13) and (14) condition above, and substituting them into condition 2, above, we see that

$$-e^{-\lambda'_{13}} = -e^{-\lambda_{13}} \text{ and } -e^{-\lambda'_{23}} = -e^{-\lambda'_{23}}. \quad (15)$$

Equation (15) is true for arbitrary operating points if and only if $(\lambda_{13}, \lambda_{23}) = (\lambda'_{13}, \lambda'_{23})$. Given condition 1, above, we have $\mathrm{T3F}(\lambda_{13}, \lambda_{23}) = \mathrm{T3F}'(\lambda_{13}, \lambda_{23})$. Note that, using the class labeling in this work, T3F and T3F′ are cumulative density functions (CDF), i.e.,

$$\mathrm{T3F}(\lambda_{13}, \lambda_{23}) = \int_{-\infty}^{\lambda_{13}} \int_{-\infty}^{\lambda_{23}} \mathrm{p}_3(x, y)\, dxdy \quad (16)$$

where $p_3()$ is the joint distribution of the log likelihood ratios λ_{13} and λ_{23} under the Class 3 hypothesis. Since the two distributions always have identical CDFs at the same $(\lambda_{13}, \lambda_{23})$ location, then the two distributions are identical. Thus the two systems have identical Class 3 likelihood ratio distributions. Using (12), we see that the likelihood ratio distributions of Classes 1 and 2 can be determined by that of Class 3. As a result, the two systems have identical Class 1 and 2 likelihood ratio distributions, respectively. This proves that the only way two systems can have identical 2-D ROC surfaces is if they have identical likelihood ratio distributions.

C. Theorem of Uniqueness and Corollary on ROC Analysis

Given the unique relationship between a three-class ROC surface and the triplet of likelihood ratio distributions, we arrive at the following theorem.

Theorem of Uniqueness—Consider two classification systems based on likelihood ratios. Let each system have continuous and differentiable 2-D sensitivity-space ROC surfaces. The systems have the same triplet of likelihood ratio (hypothesis-conditional) distributions if and only if they have the same 2-D sensitivity-space ROC surfaces.

In addition, the above theorem states that there is a one-to-one relationship between the 2-D ROC surface and the triplet of likelihood ratio distributions, assuming such a triplet exists. The intersection of the 2-D surface with the three planes defined by pairs of axes produces the three ROC curves, one for each of the three combinations of the classes taken 2 at a time. As a result, the ROC curve must uniquely determine the pair-wise likelihood ratio distributions. We thus have the following corollary:

Corollary—Consider two classification systems whose pairs of likelihood ratios are known. Let each system have continuous and differentiable ROC curves. The systems have the same pair of likelihood ratio distributions if and only if they have the same ROC curve.

Note that the Corollary can be proved directly from the properties of likelihood ratios in the binary classification, i.e., (11). We leave the proof to the reader.

V. Conclusion

Previously, we have shown that the proposed three-class ROC analysis method extends and unifies the decision theoretic (under the EEU assumption), LDA, and probabilistic foundations of ROC analysis in a three-class paradigm. In this work, we further explored the decision theoretic foundation of the proposed three-class ROC analysis method, and have built a foundation for investigating whether its generality under decision theory is restricted by the EEU assumption or not.

In particular, we have proved that two systems have the same triplet of likelihood ratio (hypothesis-conditional) distributions if and only if they have the same 2-D sensitivity-space ROC surfaces (assuming both systems have continuous and differentiable 2-D sensitivity-space ROC surfaces). Thus, the 2-D sensitivity-space ROC surface generated from likelihood ratio distributions uniquely characterizes the complete optimal three-class task performance descriptor—the triplet of likelihood ratio distributions. In other words, contrary to previous understanding, the 2-D sensitivity-space ROC provides complete information about three-class task performance of the likelihood ratio classifier.

Note that the proof of this is for the analytic 2-D ROC surface generated from the likelihood ratio distributions, and not empirical observer data. This result suggests that the 2D ROC surface may have an important role in developing a practical method for ranking of diagnostic systems in terms of three-class classification performance. However, significant theoretical work remains before the implications of the result of this paper are fully practical.

Acknowledgments

The authors would like to thank Dr. C. Metz from the University of Chicago, Dr. F. Samuelson and his group from FDA, Dr. E. Clarkson from the University of Arizona, Dr. G. Parmi-giani and his group, Dr. S. Chen, Dr. B. Caffo, and Dr. Jingyan Xu from Johns Hopkins University for their thought provoking discussions related to this work.

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

Appendix I

Now we derive the partial derivatives of TiF ($i = 1, 2, 3$) with respect to λ_{13} and λ_{23}.

A. Geometric Derivation of the Partial Derivative of T1F With Respect to λ_{13} and λ_{23}

We first compute the partial derivative of $T1F$ with respect to the two log likelihood ratios. As shown in Fig. 5(a), on the log likelihood decision space, for a decision structure centered on (λ_{13}, λ_{23}), T1F(λ_{13}, λ_{23}) is an integral of $p_1(\lambda_{13}, \lambda_{23})$ over the area for Class 1 decision, i.e., the shaded area in Fig. 5(a), where $p_i(\lambda_{13}, \lambda_{23})$ ($i = 1, 2, 3$) is the log likelihood ratio distribution under ith hypothesis.

Note that the partial derivative is defined as derivatives of a function of multiple variables when all but the variable of interest are held fixed during the computation of the derivative, and the derivative is defined as an infinitesimal change in the function with respect to one of its variables. In Fig. 5(b) and (c), we illustrate how T1F(λ_{13}, λ_{23}) changes with respect to a small change of λ_{13} and λ_{23}, respectively. In particular, Fig. 5(b) shows the computation of ∂T1F(λ_{13}, λ_{23})/$\partial\lambda_{13}$. When the decision structure moves an infinitesimal $\Delta\lambda_{13}$ along the

λ_{13} axis (solid line decision structure → dashed line decision structure), we see that T1F(λ_{13}, λ_{23}) decreases, and the amount of decrease is equivalent to integrating $p_1(\lambda_{13}, \lambda_{23})$ along the small strip as shown in the shaded area in Fig. 5(b). This observation can be written down mathematically. Given the definition of the partial derivative

$$\frac{\partial T1F}{\partial \lambda_{13}} = \lim_{\Delta\lambda_{13}\to 0} \frac{T1F(\lambda_{13}+\Delta\lambda_{13}, \lambda_{23}) - T1F(\lambda_{13}, \lambda_{23})}{\Delta\lambda_{13}}. \quad \text{(A1)}$$

As in Fig. 5(b), when $\Delta\lambda_{13} \to 0$, the difference between $T1F(\lambda_{13} + \Delta\lambda_{13}, \lambda_{23})$ and $T1F(\lambda_{13}, \lambda_{23})$ is expressed as integration of $p_1(\lambda_{13}, \lambda_{23})$ along the vertical line and the 45° multiplied by $\Delta\lambda_{13}$. Consequently, the partial derivative is

$$\frac{\partial T1F}{\partial \lambda_{13}} = -\left[\int_{-\infty}^{\lambda_{23}} p_1(\lambda_{13}, y)\, dy + \int_{\lambda_{23}}^{-\infty} p_1(\lambda_{13}+y-\lambda_{23}, y)\, dy\right]. \quad \text{(A2)}$$

The negative sign is because that $T1F(\lambda_{13} + \Delta\lambda_{13}, \lambda_{23}) < T1F(\lambda_{13}, \lambda_{23})$ for $\Delta\lambda_{13} > 0$.

Similarly, the partial derivative of $T1F$ with respect to λ_{23} is equivalent to integrating $p_1(\lambda_{13}, \lambda_{23})$ along the 45° line, as shown in Fig. 5(c), i.e.,

$$\frac{\partial T1F}{\partial \lambda_{23}} = \int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx. \quad \text{(A3)}$$

B. More Rigorous Computation of Partial Derivative of T1F With Respect to λ_{13} and λ_{23}

Observing Fig. 5(a) shows that the area for Class 1 decision is bounded by a vertical ray extending from the center of the decision structure to $-\infty$ and a second ray from the center of the decision structure along the 45° line from the center of the decision structure toward (∞,∞). Thus, T1F(λ_{13}, λ_{23}) can be expressed as

$$T1F(\lambda_{13}, \lambda_{23}) = \int_{\lambda_{13}}^{\infty}\int_{\lambda_{23}}^{x-\lambda_{13}+\lambda_{23}} p_1(x, y)\, dy dx + \int_{\lambda_{13}}^{\infty}\int_{-\infty}^{\lambda_{23}} p_1(x, y)\, dy dx \quad \text{(A4)}$$

where the first term is an integral of $p_1()$ over the shaded region above the dashed line in Fig. 5(a), and the second term is an integral of p_1 over the shaded region below the dashed line in Fig. 5(a). The partial derivative of T1F with respect to λ_{13} can be easily computed using (A4) by applying the Leibnitz's rule to the differentiation of the first term. We leave the derivation of T1F with respect to λ_{23} to the readers.

C. Partial Derivatives of T2F and T3F With Respect to λ_{13} and λ_{23}

Using the methods in the previous section, we can easily compute the partial derivatives of $T2F$ and $T3F$ with respect to λ_{13} and λ_{23}. The illustrations for these partial derivatives are given in Fig. 6. For more rigorous derivation, please refer to our previous publication [34]. In particular, a change of $\Delta\lambda_{13} > 0$ results in an increase of $T2F$ in the shaded area as illustrated in Fig. 6(a). When $\Delta\lambda_{13} \to 0$, $\partial T2F/\partial\lambda_{13}$ equals $p_2(\lambda_{13}, \lambda_{23})$ integrated along the 45° line. Similarly, $\partial T2F/\partial\lambda_{23}$ is $p_2(\lambda_{13}, \lambda_{23})$ integrated long the 45° and the horizontal line [Fig. 6(b)]; $\partial T3F/\partial\lambda_{13}$ is $p_3(\lambda_{13}, \lambda_{23})$ is integrated along the vertical line [Fig. 6(c)];

and $\partial T3F/\partial\lambda_{23}$ is $p_3(\lambda_{13}, \lambda_{23})$integrated along the horizontal line. These partial derivatives are expressed as

$$\frac{\partial T2F}{\partial\lambda_{13}} = \int_{\lambda_{13}}^{\infty} p_2(x, x+\lambda_{23}-\lambda_{13})\,dx \quad \text{(A5)}$$

$$\frac{\partial T2F}{\partial\lambda_{23}} = -\int_{-\infty}^{\lambda_{13}} p_2(x, \lambda_{23})\,dx - \int_{\lambda_{13}}^{\infty} p_2(x, x+\lambda_{23}-\lambda_{13})\,dx \quad \text{(A6)}$$

$$\frac{\partial T3F}{\partial\lambda_{13}} = -\int_{-\infty}^{\lambda_{23}} p_3(\lambda_{13}, y)\,dy \quad \text{(A7)}$$

and

$$\frac{\partial T3F}{\partial\lambda_{23}} = \int_{-\infty}^{\lambda_{13}} p_3(x, \lambda_{23})\,dx. \quad \text{(A8)}$$

Appendix II

In this Appendix, we provide a method for deriving the partial derivatives of the 2-D ROC surface, e.g.,$\partial T2F/\partial T1F$. We use $p_i(\lambda_{13}, \lambda_{23})$ ($i = 1, 2, 3$) to denote the log likelihood ratio distributions for Class i and $g_i(\Lambda_{13},\Lambda_{23})$ the corresponding likelihood ratio distribution for Class i. By definition of the probability density functions, we have

$$Jg_i(\Lambda_{13}, \Lambda_{23}) = Jg_i\left(e^{\lambda_{13}}, e^{\lambda_{23}}\right) = p_i(\lambda_{13}, \lambda_{23}) \quad \text{(A9)}$$

where $J = (\partial(\Lambda_{13}, \Lambda_{23})/\partial(\lambda_{13}, \lambda_{23})) = e^{(\lambda_{13}+\lambda_{23})} = J(\lambda_{13},\lambda_{23})$.

In (A9), J is the Jacobian.

A. Geometric Derivation of ∂T2F/∂T1F

Given the intuitive description of TiF with respect to λ_{13} and λ_{23}, we provide a intuitive description of $dTiF$ ($i = 1, 2, 3$), shown in Fig. 7. The total derivative of dTiF ($i = 1, 2, 3$), which is expressed as

$$dTiF = \frac{\partial TiF}{\partial\lambda_{13}}d\lambda_{13} + \frac{\partial TiF}{\partial\lambda_{23}}d\lambda_{23}. \quad \text{(A10)}$$

We thus see that $dTiF$ ($i = 1, 2, 3$) is a sum of infinitesimal change of TiF along both λ_{13} and λ_{23} directions, i.e., $(\partial TiF/\partial\lambda_{13})\Delta\lambda_{13}$ and $(\partial TiF/\partial\lambda_{23})\Delta\lambda_{23}$. We have illustrated the derivations of $\partial TiF/\partial\lambda_{13}$ and $\partial TiF/\partial\lambda_{23}$ in the previous sections, Fig. 7 illustrates the derivation of $\partial T2F/\partial T1F$. Again, the partial derivative is defined as derivatives of a function of multiple variables when all but the variable of interest are held fixed during the differentiation. Partial derivative $\partial T2F/\partial T1F$ is equivalent to $dT2F/dT1F$ given $dT3F = 0$. Shown in Fig. 7, moving the decision structure from the original location (represented by solid lines) to another location (represented by the dashed line), i.e., a shift of $(\Delta\lambda_{13}, \Delta\lambda_{23})$ where $\Delta\lambda_{13} \to 0$ and $\Delta\lambda_{23} \to 0$, results in small change in all three sensitivities. Fig. 7(c) shows the change of $T3F$. Shifts of decision structure to the right increase $T3F$, while shifts downward decrease $T3F$. In order for $dT3F = 0$, we see that the increase of $T3F$ along λ_{13}

direction must equal the decrease of $T3F$ along λ_{23} direction. Fig. 7(a) shows the change of $T1F$, where the shift of decision structure results in decrease of $T1F$. As a result, $dT1F$ is a sum of the change in $T1F$ caused by $\Delta\lambda_{13}$ and the change in $T1F$ caused by $\Delta\lambda_{23}$. Fig. 7(b) shows the change of $T2F$, where the shift of decision structure results in increase of $T2F$.

B. Computation of ∂T2F/∂T1F

The partial derivative is defined as derivatives of a function of multiple variables when all but the variable of interest are held fixed during the differentiation. Using $\partial T2F/\partial T1F$ as an example, in the sensitivity space that is spanned by $T1F$, $T2F$ and $T3F$, the 2-D ROC surface can be expressed as a function of the three sensitivities, i.e., $T2F$ = T2F($T1F$, $T3F$). Thus, $\partial T2F/\partial T1F$ is obtained by holding $T3F$ fixed during the differentiation, i.e.,

$$\frac{\partial T2F}{\partial T1F} = \frac{dT2F}{dT1F} \text{ given } dT3F = 0 \quad \text{(A11)}$$

where $dTiF$ (i = 1, 2, 3) is the total derivative of Class i. Equation (A11) shows that the key to computing the partial derivative $\partial T2F/\partial T1F$ of is to compute the total derivatives $dTiF$ (i = 1, 2, 3), as expressed in (A10), which shows that $dTiF$ is determined by the partial derivative of TiF (i = 1, 2, 3) with respect to λ_{13} and λ_{23}. In Appendix I, we have provided derivations of $\partial TiF/\partial\lambda_{13}$ and $\partial TiF/\partial\lambda_{23}$.

Now our goal is to substitute $\partial TiF/\partial\lambda_{13}$ and $\partial TiF/\partial\lambda_{23}$, derived in Appendix I, into (A10) and to simplify it. The key to the simplification is the property of likelihood ratios. Using the properties of likelihood hood ratios in (12), we have

$$e^{\lambda_{13}} = \frac{g_1(e^{\lambda_{13}}, e^{\lambda_{23}})}{g_3(e^{\lambda_{13}}, e^{\lambda_{23}})}$$
$$e^{\lambda_{23}} = \frac{g_2(e^{\lambda_{13}}, e^{\lambda_{23}})}{g_3(e^{\lambda_{13}}, e^{\lambda_{23}})}$$

and

$$e^{\lambda_{23}-\lambda_{13}} = \frac{g_2\left(e^{\lambda_{13}} e^{\lambda_{23}}\right)}{g_1\left(e^{\lambda_{13}}, e^{\lambda_{23}}\right)}. \quad \text{(A12)}$$

Apply the Jacobian for the log transformations in (A9) to (A12), we have

$$e^{\lambda_{13}} = \frac{p_1(\lambda_{13}, \lambda_{23})}{p_3(\lambda_{13}, \lambda_{23})}$$
$$e^{\lambda_{23}} = \frac{p_2(\lambda_{13}, \lambda_{23})}{p_3(\lambda_{13}, \lambda_{23})}$$

and

$$e^{\lambda_{23}-\lambda_{13}} = \frac{p_2\left(\lambda_{13}, \lambda_{23}\right)}{p_1\left(\lambda_{13}, \lambda_{23}\right)}. \quad \text{(A13)}$$

To compute $\partial T2F/\partial T1F$, we use (A13) to rewrite $\partial TiF/\partial\lambda_{13}$ and $\partial TiF/\partial\lambda_{23}$, and obtain new expressions for $dT1F$ and $dT2F$. The goal of doing so is to cancel out some of the terms when computing the division of $dT2F$ by $dT1F$.

We first compute $dT1F$, i.e., the denominator. Substituting (A13) into (A12), we have

$$\frac{\partial T1F}{\partial \lambda_{13}} = -e^{\lambda_{13}} \int_{-\infty}^{\lambda_{23}} p_3(\lambda_{13}, y)\, dy - \int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx. \quad \text{(A14)}$$

Similarly, $\partial T1F/\partial\lambda_{23}$ is expressed as in (A3). Substituting (A13) and (A3) into (A10) and rearranging the terms, we have

$$dT1F = -e^{\lambda_{13}} \left[\left(\int_{-\infty}^{\lambda_{23}} p_3(\lambda_{13}, y)\, dy \right) d\lambda_{13} \right] + \left(\int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx \right) (d\lambda_{23} - d\lambda_{13}). \quad \text{(A15)}$$

We then compute $dT2F$. Substituting the third equation in (A13) into (A5), $\partial T2F/\partial\lambda_{13}$ is expressed as

$$\frac{\partial T2F}{\partial \lambda_{13}} = -e^{\lambda_{23}-\lambda_{13}} \int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx. \quad \text{(A16)}$$

Similarly, applying the properties of likelihood ratios and the Jacobian for change of variables as in (A13), (A6) can be expressed as

$$\frac{\partial T2F}{\partial \lambda_{23}} = -e^{\lambda_{23}} \int_{-\infty}^{\lambda_{13}} p_3(x, \lambda_{23})\, dx - e^{\lambda_{23}-\lambda_{13}} \int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx. \quad \text{(A17)}$$

Thus, $dT2F$ can be obtained by substituting (A16) and (A17) into (A10), and rearrange the terms, i.e.,

$$\begin{aligned} dT2F = &-e^{\lambda_{23}} \left[\left(\int_{-\infty}^{\lambda_{13}} p_3(x, \lambda_{23})\, dx \right) d\lambda_{23} \right] \\ &-e^{\lambda_{23}-\lambda_{13}} \left(\int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx \right) \\ &\times (d\lambda_{23} - d\lambda_{13}). \end{aligned} \quad \text{(A18)}$$

Substituting (A15) and (A18) into (A11), we have

$$\frac{\partial T2F}{\partial T1F} = \frac{A}{B}$$

where

$$\begin{aligned} A &= e^{\lambda_{23}} \left[\left(-\int_{-\infty}^{\lambda_{13}} p_3(x, \lambda_{23})\, dx \right) d\lambda_{23} \right] \\ &\quad - e^{\lambda_{23}-\lambda_{13}} \left(\int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx \right) \\ &\quad \times (d\lambda_{23} - d\lambda_{13}) \\ B &= -e^{\lambda_{13}} \left[\left(\int_{-\infty}^{\lambda_{23}} p_3(\lambda_{13}, y)\, dy \right) d\lambda_{13} \right] \\ &\quad + \left(\int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx \right) \\ &\quad \times (d\lambda_{23} - d\lambda_{13}). \end{aligned} \tag{A19}$$

To further simplify (A19), we substitute the expressions for $\partial T3F/\partial\lambda_{13}$ derived in (A7) and $\partial T3F/\partial\lambda_{23}$ derived in (A8) into $dT3F = 0$. We see that

$$\left(\int_{-\infty}^{\lambda_{23}} p_3(\lambda_{13}, y)\, dy \right) d\lambda_{13} = -\left(\int_{-\infty}^{\lambda_{13}} p_3(x, \lambda_{23})\, dx \right) d\lambda_{23}. \tag{A20}$$

Equation (A20) shows that the terms inside the square brackets in the denominator and numerator of (A19) are equal. Thus, (A19) becomes

$$\frac{\partial T2F}{\partial T1F} = \frac{A}{B}$$

where

$$\begin{aligned} A &= e^{\lambda_{23}} \left[\left(\int_{-\infty}^{\lambda_{23}} p_3(\lambda_{13}, y)\, dy \right) d\lambda_{13} \right] \\ &\quad - e^{\lambda_{23}-\lambda_{13}} \left(\int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx \right) \\ &\quad \times (d\lambda_{23} - d\lambda_{13}) \\ B &= -e^{\lambda_{13}} \left[\left(\int_{-\infty}^{\lambda_{23}} p_3(\lambda_{13}, y)\, dy \right) d\lambda_{13} \right] \\ &\quad + \left(\int_{\lambda_{13}}^{\infty} p_1(x, x+\lambda_{23}-\lambda_{13})\, dx \right) \\ &\quad \times (d\lambda_{23} - d\lambda_{13}). \end{aligned} \tag{A21}$$

Factoring $-e^{\lambda_{23}-\lambda_{13}}$ from the numerator of (A21), we see that the remaining terms are equal, leaving

$$\frac{\partial T2F}{\partial T1F} = -e^{\lambda_{23}-\lambda_{13}} = -\Lambda_{12}^{-1}. \tag{A22}$$

Using the same approach, we can derive

$$\frac{\partial T3F}{\partial T2F} = -e^{-\lambda_{23}} = -\Lambda_{23}^{-1}$$

and

$$\frac{\partial T3F}{\partial T1F} = -e^{-\lambda_{13}} = -\Lambda_{13}^{-1}. \quad \text{(A23)}$$

References

1. 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.
2. Swets, JA.; Pickett, RM. Evaluation of Diagnostic System: Methods From Signal Detection Theory. New York: Academic; 1982.
3. Yeung KY, Bumgarner RE, Raftery AE. Bayesian model averaging: Development of an improved multi-class, gene selection and classification tool for microarray data. Bioinformatics. May 15.2005 21:2394–2402. [PubMed: 15713736]
4. Xiong CJ, van Belle G, Miller JP, Morris JC. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. Stat Med. Apr.2006 25:1251–1273. [PubMed: 16345029]
5. Scurfield BK. Generalization of the theory of signal detectability to n-event m-dimensional forced-choice tasks. J Math Psychol. Mar.1998 42:5–31. [PubMed: 9606159]
6. Scurfield BK. Multiple-event forced-choice tasks in the theory of signal detectability. J Math Psychol. Sep.1996 40:253–269. [PubMed: 8979976]
7. Sahiner B, Chan H-P, Hadjiiski LM. Performance analysis of 3-class classifiers: Properties of the 3D ROC surface and the normalized volume under the surface. Proc SPIE. 2006; 6146
8. Sahiner B, Chan HP, Hadjilski 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.
9. Patel AC, Markey MK. Comparison of three-class classification performance metrics: A case study in breast cancer CAD. Proc SPIE. 2005
10. Nakas CT, Yiannoutsos CT. Ordered multiple-class ROC analysis with continuous measurements. Stat Med. Nov.2004 23:3437–3449. [PubMed: 15505886]
11. Mossman D. Three-way ROCs. Med Decision Making. Jan-Mar;1999 19:78–89.
12. Metz, CE. The optimal decision variable: Likelihood ratio Univ. Chicago. Chicago, IL: Mathematics for medical physicists handout; 2000.
13. Kijewski MF, Swensson RG, Judy PF. Analysis of rating data from multiple-alternative tasks. J Math Psychol. 1989; 33:428–451.
14. 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 Diag Lab Immunol. May.2005 12:566–574.
15. Heckerling PS. Parametric three-way receiver operating characteristic surface analysis using mathematica. Med Decis Making. 2001; 21:409–417. [PubMed: 11575490]
16. 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.
17. Fukunaga, K. Introduction to Statistical Pattern Recognition. New York: Academic; 1990.
18. Everson RM, Fieldsend JE. Multi-class ROC analysis from a multi-objective optimisation perspective. Pattern Recognit Lett. Jun.2006 27:918–927.
19. 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.
20. 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.
21. Edwards, DC.; Metz, CE. Optimality of a utility-based performance metric for ROC analysis. In: Sahiner, B.; Manning, DJ., editors. Proc SPIE Med Imag 2008: Image Perception, Observer Performance, Technology Assessment. Vol. 6917. 2008.

22. Edwards DC, Metz CE. A utility-based performance metric for ROC analysis of N-class classification tasks. Proc SPIE Med Imag 2007: Image Perception, Observer Performance, Technol Assessment. 2007:6515031–65150310.
23. Edwards DC, Metz CE. Optimization of restricted ROC surfaces in three-class classification tasks. IEEE Trans Med Imag. Oct.2007 26(10):1345–1356.
24. 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.
25. 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.
26. 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. Jan.2004 31:81–90. [PubMed: 14761024]
27. Edwards, DC. Ideal Observer Estimation and Generalized ROC Analysis for Computer-Aided Diagnosis Univ Chicago. Chicago, IL: 2003.
28. Dreiseitl S, Ohno-Machado L, Binder M. Comparing three-class diagnostic tests by three-way ROC analysis. Med Decis Making. Jul-Sep;2000 20:323–31. [PubMed: 10929855]
29. Chan H-P, 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.
30. Van Trees, HL. Detection, Estimation and Modulation Theory: Part I. New York: Wiley; 1968.
31. 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.
32. 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.
33. He X, Frey EC. An optimal three-class linear observer derived from decision theory. IEEE Trans Med Imag. Jan.2007 26(1):77–83.
34. 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.
35. Barrett, HH.; Myers, KJ. Foundations of Image Science. New York: Wiley; 2003.
36. Edwards DC, Metz CE. Evaluating Bayesian ANN estimates of ideal observer decision variable by comparison with identity functions. Proc SPIE. 2005; 5749:174–182.

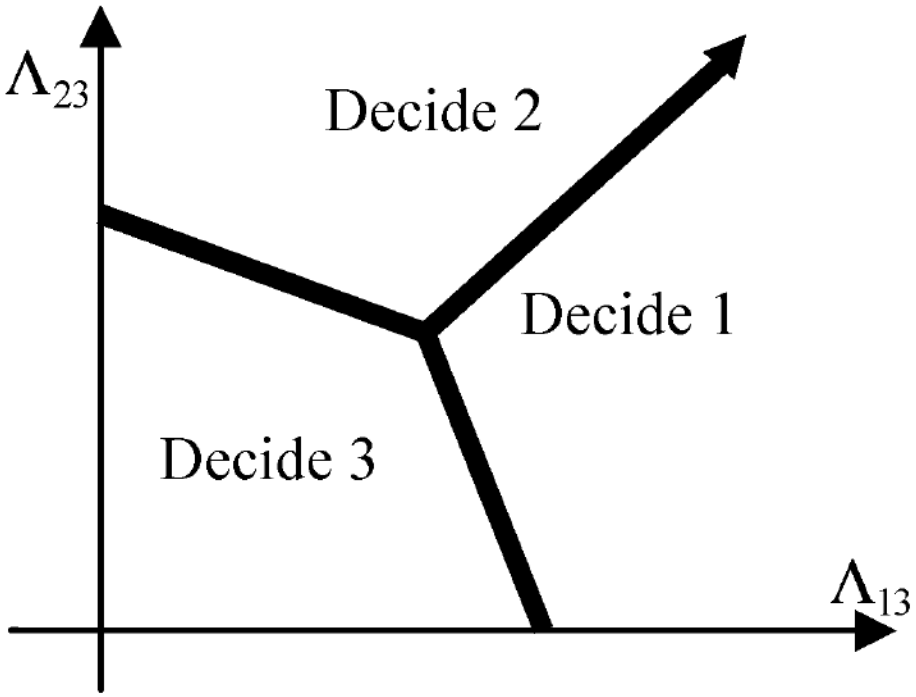

Fig. 1.
Three-class decision plane, where likelihood ratios Λ_{13} and Λ_{23} are used as the decision variables. The decision rules, represented by the decision structure, have 5 degrees-of-freedom.

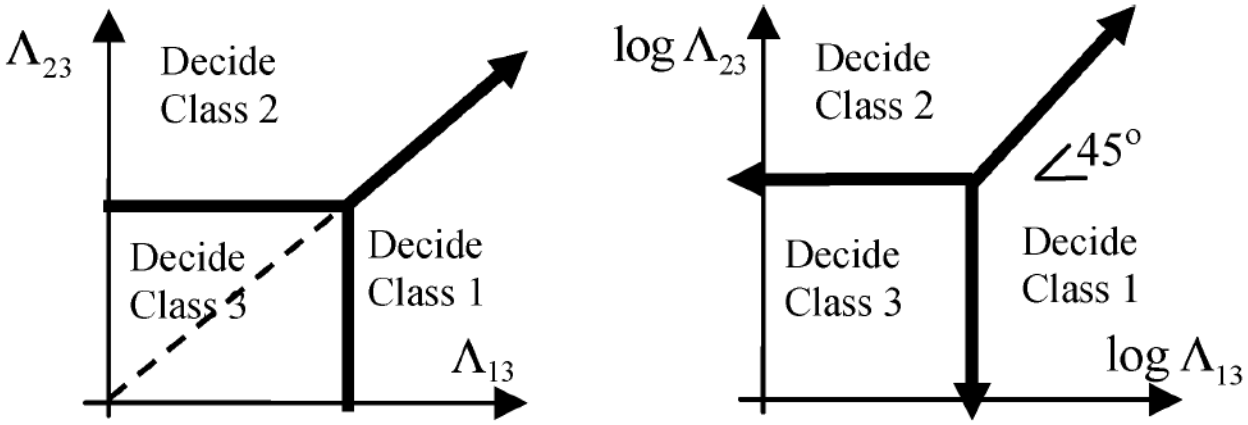

Fig. 2.
Three-class likelihood ratio and log likelihood ratio decision plane under EEU assumption.

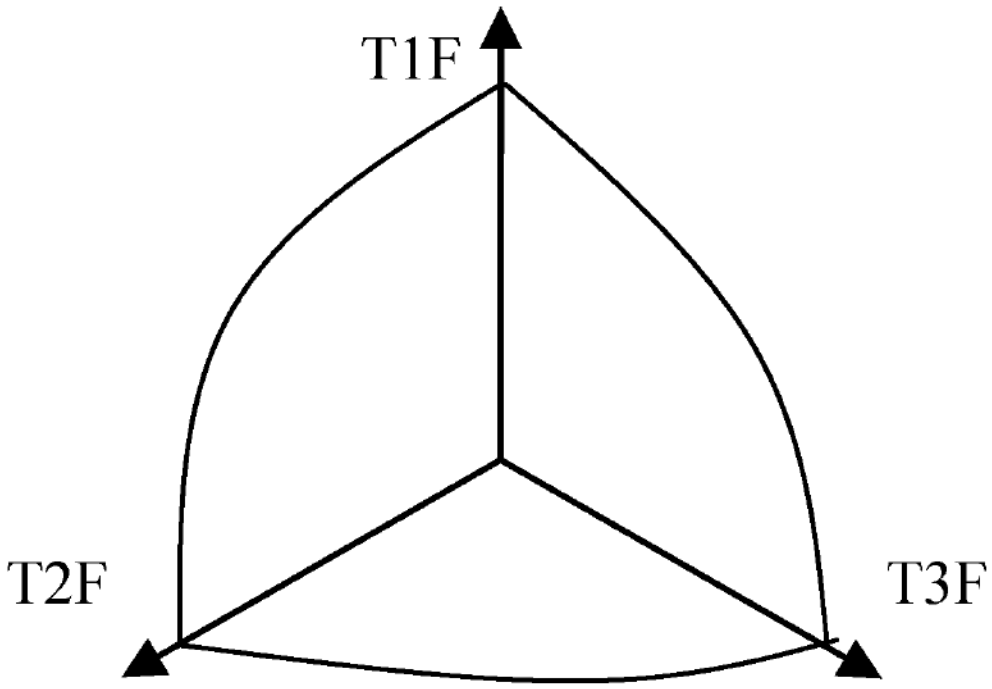

Fig. 3.
Example of the 2-D practical three-class ROC surface in the 3-D sensitivity space.

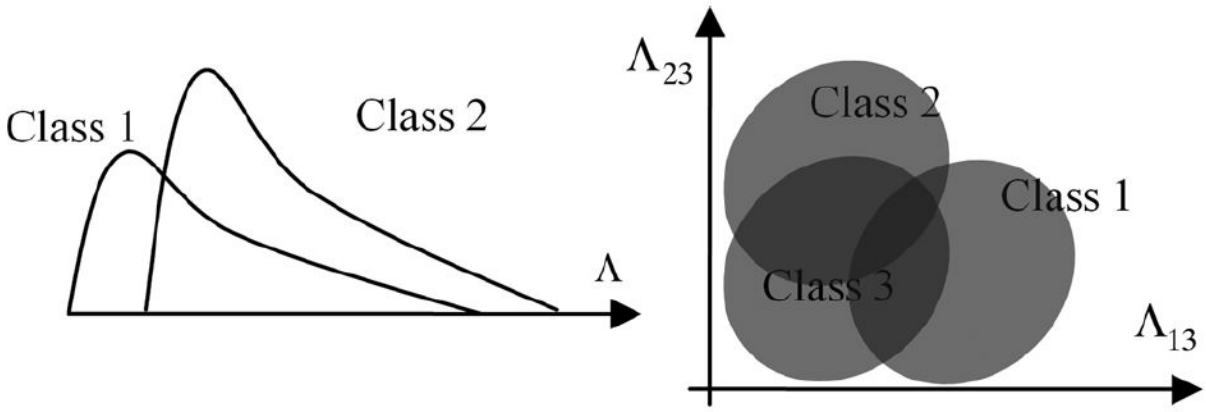

Fig. 4.
Examples of a pair of likelihood ratio distributions in a binary decision space and a triplet of likelihood distributions in a three-class decision space.

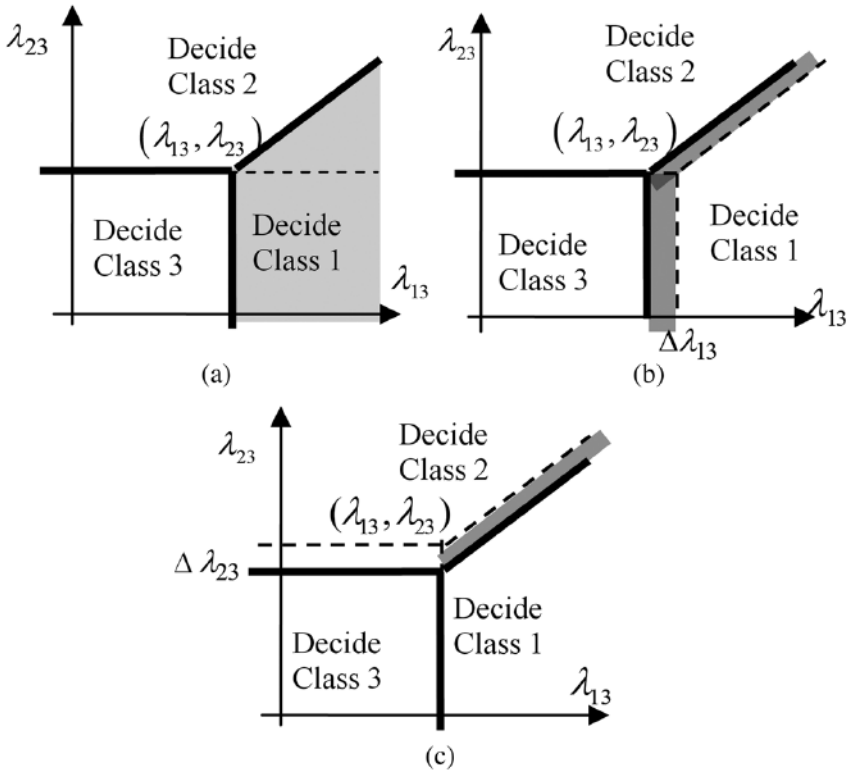

Fig. 5.

Illustration for the computation of $T1F$ and its partial derivatives. (a) $T1F$ is computed by integrating $p_1(\lambda_{13}, \lambda_{23})$ over the shaded region. (b) Moving the decision structure to $(\lambda_{13} + \Delta\lambda_{13}, \lambda_{23})$, $\partial T1F/\partial\lambda_{13}$ is the integral of $p_1(\lambda_{13}, \lambda_{23})$ over the shaded stripe when $\Delta\lambda_{13} \rightarrow 0$, i.e., the integral of $p_1(\lambda_{13}, \lambda_{23})$ along the vertical and 45-degree line. Since $T1F$ is decreasing, this partial derivative is negative. (c) Moving the decision structure to $(\lambda_{13}, \lambda_{23} + \Delta\lambda_{23})$, $\partial T1F/\partial\lambda_{23}$, is the integral of $p_1(\lambda_{13}, \lambda_{23})$ over the shaded stripe when $\Delta\lambda_{23} \rightarrow 0$, i.e., the integral of $p_1(\lambda_{13}, \lambda_{23})$ along 45° line.

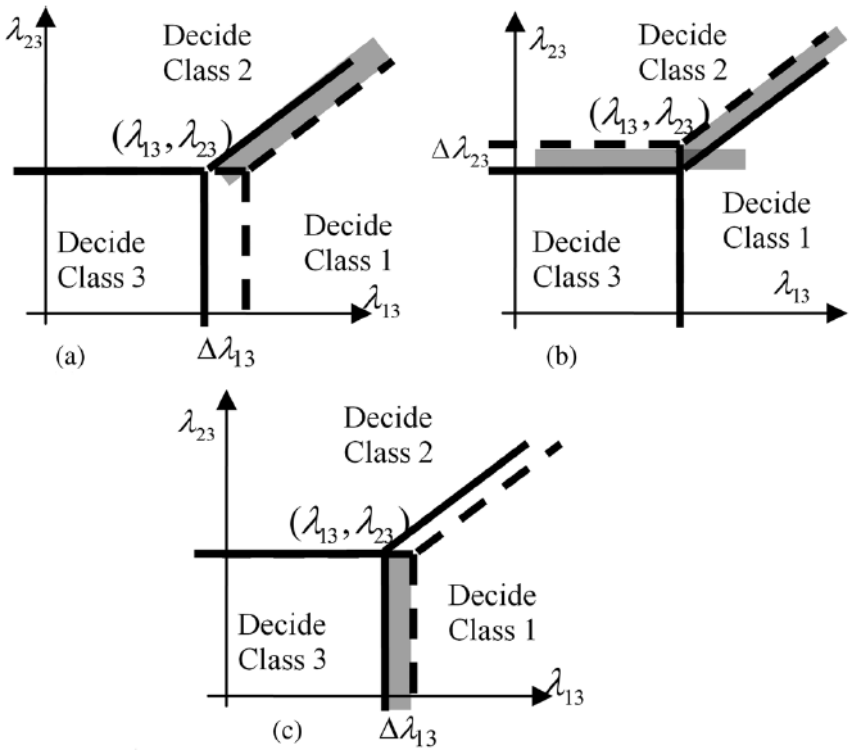

Fig. 6.

Illustration of the derivations of the partial derivative of $T2F$ and $T3F$ with respect to λ_{13} and λ_{23}. (a) Illustration of $\partial T2F/\partial\lambda_{13}$; (b) illustration of $\partial T2F/\partial\lambda_{23}$; and (c) illustration of $\partial T3F/\partial\lambda_{13}$.

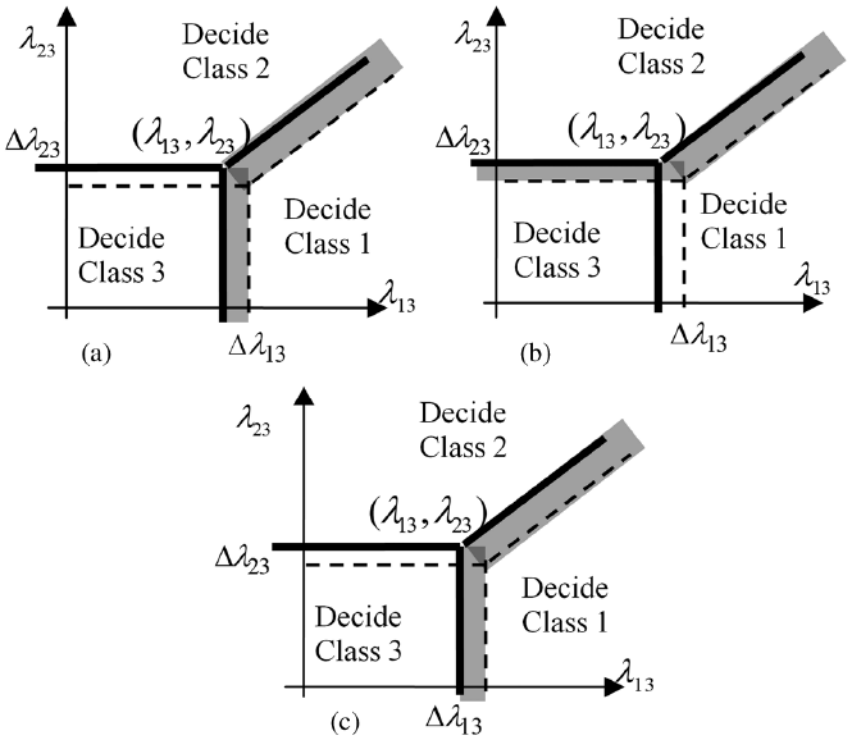

Fig. 7.

Illustration for the computation of $\partial T2F/\partial T1F$. (a) Moving the decision structure while keeping $T3F$ constant; (b) the increase of $T3F$ after moving the decision structure; and (c) the decrease of $T1F$ after moving the decision structure.

TABLE I

Decision Table For Three-Class Classification

		True Status of an object		
		Class 1	**Class 2**	**Class 3**
Decision	Class 1	T1 *	F12	F13
	Class 2	F21 *	T2	F23
	Class 3	F31	F32	T3

* T*i* (*i* = 1, 2, 3) True decision, i.e., Class *i* decision when the underlying truth is also Class *i*.

* F*ij* (*i, j* = 1, 2, 3 *i* ≠ *j*) False decision, i.e., Class *i* decision when the underlying truth is Class *j*.