Review of several proposed three-class classification decision rules and their relation to the ideal observer decision rule

Darrin C. Edwards* and Charles E. Metz

Department of Radiology, The University of Chicago, Chicago, IL 60637

ABSTRACT

We analyzed a variety of recently proposed decision rules for three-class classification from the point of view of ideal observer decision theory. We considered three-class decision rules which have been proposed recently: one by Scurfield, one by Chan *et al.*, and one by Mossman. Scurfield's decision rule can be shown to be a special case of the three-class ideal observer decision rule in two different situations: when the pair of decision variables is the pair of likelihood ratios used by the ideal observer, and when the pair of decision variables is the pair of logarithms of the likelihood ratios. Chan *et al.* start with an ideal observer model, where two of the decision lines used by the ideal observer overlap, and the third line becomes undefined. Finally, we showed that the Mossman decision rule (in which a single decision line separates one class from the other two, while a second line separates those two classes) cannot be a special case of the ideal observer decision rule. Despite the considerable difficulties presented by the three-class classification task compared with two-class classification, we found that the three-class ideal observer provides a useful framework for analyzing a wide variety of three-class decision strategies.

Keywords: ROC analysis, three-class classification, ideal observer decision rules

1. INTRODUCTION

We are attempting to develop a fully automated mass lesion classification scheme for computer-aided diagnosis (CAD) in mammography. This scheme will combine two schemes developed at the University of Chicago: one for automatically detecting mass lesions in mammograms,[1–5] and one for classifying known lesions as malignant or benign.[6–10] Combining these two types of CAD scheme is inherently difficult, because the output of the detection scheme will necessarily include false-positive (FP) computer detections in addition to the malignant and benign lesions to be classified. These FP computer detections correspond to objects which were by design not included in the training sample of the classification scheme, because they are not members of the data population (benign and malignant mass breast lesions) for which the classification scheme was created. It is clear then that the detection scheme's output cannot be used unmodified as the input to the classification scheme.

Our approach has been to treat this problem explicitly as a three-class classification task. That is, the outputs of the detection scheme should be classified as malignant lesions, benign lesions, and non-lesions (FP computer detections), and the classifier to be estimated is the ideal observer decision rule for this task. Such an approach presents considerable difficulties of its own. On the one hand, decision rules, in particular ideal observer decision rules, increase rapidly in complexity with the number of classes involved. On the other hand, a fully general performance evaluation method, such as a three-class extension of receiver operating characteristic (ROC) analysis, has yet to be developed.

The explicit form of the ideal observer in a three-class classification task has been known for some time.[11] For the reasons just stated, however, a practical method for estimating and evaluating observer performance based on an ideal observer model has proven elusive, despite the success of the two-class binormal ideal observer model.[12] Nevertheless, pragmatic observer decision rule models for three-class classification tasks have been proposed relatively recently by several groups of researchers. In some cases, these models are motivated more by considerations of tractability than of complete generality. This is of course understandable given the inherent difficulties of three-class classification; however, we thought it might be of interest to analyze a number of recently proposed three-class decision rule models within an ideal observer decision rule framework.

*Correspondence: E-mail: d-edwards@uchicago.edu; Telephone: 773 834 5094; Fax: 773 702 0371

In the next section, we review the three-class ideal observer decision rule. In the following three sections, we review recently proposed three-class decision rule models: one by Scurfield,[13] one by Chan *et al.*,[14] and one by Mossman.[15] In each case, the given decision rule is analyzed in terms of the ideal observer decision rule; where necessary or expedient, assumptions are made about the observer's decision variables in order to facilitate this analysis. We emphasize that we do not attempt a review of the experimental methods in the works discussed; we are specifically interested only in the form of the decision rule which serves as the starting point for each work. The results of our analyses are briefly summarized in Sec. 6.

2. THE THREE-CLASS IDEAL OBSERVER

It can be shown[11, 16] that an N-class ideal observer makes decisions regarding statistically variable observations $\vec{\mathbf{x}}$ by partitioning a likelihood ratio decision variable space, where the boundaries of the partitions are given by hyperplanes:

$$\text{decide} \quad d = \pi_i \quad \text{iff}$$

$$\sum_{k=1}^{N-1} (U_{i|k} - U_{j|k}) P(\mathbf{t} = \pi_k) \mathrm{LR}_k \quad \geq \quad (U_{j|N} - U_{i|N}) P(\mathbf{t} = \pi_N) \qquad \{j < i\} \tag{1}$$

$$\text{and}$$

$$\sum_{k=1}^{N-1} (U_{i|k} - U_{j|k}) P(\mathbf{t} = \pi_k) \mathrm{LR}_k \quad > \quad (U_{j|N} - U_{i|N}) P(\mathbf{t} = \pi_N) \qquad \{j > i\}. \tag{2}$$

Here $U_{i|j}$ is the utility of deciding an observation is from class π_i given that it is actually from class π_j, and the $N-1$ likelihood ratios are defined as

$$\mathrm{LR}_i \equiv \frac{p_{\vec{x}}(\vec{x}|\mathbf{t} = \pi_i)}{p_{\vec{x}}(\vec{x}|\mathbf{t} = \pi_N)} \tag{3}$$

for $i < N$. We also define the actual class (the "truth") to which an observation belongs as $\mathbf{t}$, and the class to which it is assigned (the "decision") as $\mathbf{d}$, where $\mathbf{t}$ and $\mathbf{d}$ can take on any of the values $\pi_1, \ldots, \pi_i, \ldots, \pi_N$, the labels of the various classes. (We use boldface type to denote statistically variable quantities.)

The partitioning of the decision variable space is determined by the parameters

$$\gamma_{ijk} \equiv (U_{i|k} - U_{j|k}) P(\mathbf{t} = \pi_k), \tag{4}$$

with i, j, and k varying from 1 to N, and $j \neq i$. Note that these parameters are not independent, however, because

$$\gamma_{ijk} = \gamma_{kjk} - \gamma_{kik}. \tag{5}$$

We can impose the reasonable condition that the utility for correctly classifying an observation from a given class should be greater than any utility for incorrectly classifying an observation from the same class, *i.e.*, $U_{i|i} > U_{j|i} \quad \{i \neq j\}$. This gives, for $j \neq i$,

$$\gamma_{iji} > 0, \tag{6}$$

leaving $N(N-1)$ parameters (the rest are derivable from Eq. 5).

Finally, note that the hyperplanes represented by Eqs. 1 and 2 are unchanged if we multiply all of these equations by a single scalar, such as $1/(\sum_{i \neq j} \gamma_{iji})$. This leaves us with $N^2 - N - 1$ degrees of freedom, as expected.

The behavior of a three-class ideal observer is completely determined by the three decision boundary lines

$$\gamma_{121}\mathrm{LR}_1 - \gamma_{212}\mathrm{LR}_2 = \gamma_{313} - \gamma_{323} \tag{7}$$

$$\gamma_{131}\mathrm{LR}_1 + (\gamma_{232} - \gamma_{212})\mathrm{LR}_2 = \gamma_{313} \tag{8}$$

$$(\gamma_{131} - \gamma_{121})\mathrm{LR}_1 + \gamma_{232}\mathrm{LR}_2 = \gamma_{323}, \tag{9}$$

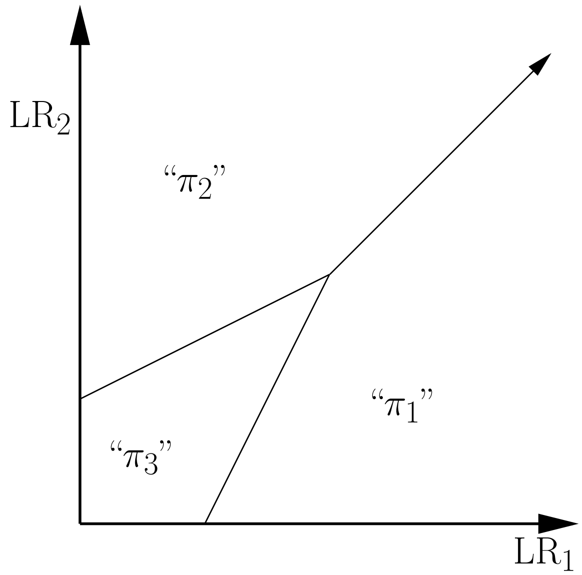

Figure 1. Example three-class ideal observer decision rule, given the values of the decision parameters $\gamma_{121} = \gamma_{212} = 3/14$ and $\gamma_{131} = \gamma_{313} = \gamma_{232} = \gamma_{323} = 1/7$. Note $\gamma_{iji} \equiv (U_{i|i} - U_{j|i})P(\mathbf{t} = \pi_k)$.

which we call, respectively, the "1-*vs.*-2" line, the "1-*vs.*-3" line, and the "2-*vs.*-3" line. Note that if any two of these lines intersect, the third line must also share this intersection point. We also emphasize the simple interpretation, from Eq. 4, of each of the γ_{iji} parameters appearing in these decision boundary line equations as the difference in utilities between a "correct" and one particular "incorrect" decision (scaled by the *a priori* probability of the true class in question); and of each difference in the γ_{iji} parameters as a difference in utilities between two possible "incorrect" decisions (again scaled by the *a priori* probability of the true class in question).

An example ideal observer decision rule for particular values of the utilities $U_{i|j}$, and hence of the parameters γ_{iji}, is shown in Fig. 1. Here we have chosen $\gamma_{121} = \gamma_{212} = 3/14$ and $\gamma_{131} = \gamma_{313} = \gamma_{232} = \gamma_{323} = 1/7$, yielding the decision boundary lines

$$\frac{3}{14}\mathrm{LR}_1 - \frac{3}{14}\mathrm{LR}_2 = 0 \quad \{\text{“1-}\textit{vs.}\text{-2”}\} \tag{10}$$

$$\frac{1}{7}\mathrm{LR}_1 - \frac{1}{14}\mathrm{LR}_2 = \frac{1}{7} \quad \{\text{“1-}\textit{vs.}\text{-3”}\} \tag{11}$$

$$-\frac{1}{14}\mathrm{LR}_1 + \frac{1}{7}\mathrm{LR}_2 = \frac{1}{7} \quad \{\text{“2-}\textit{vs.}\text{-3”}\}. \tag{12}$$

These simplify to the equations $\mathrm{LR}_2 = \mathrm{LR}_1$, $\mathrm{LR}_2 = 2\mathrm{LR}_1 - 2$, and $\mathrm{LR}_2 = \mathrm{LR}_1/2 + 1$, respectively.

3. THE SCURFIELD DECISION RULE

Scurfield investigated a decision rule applied to two-dimensional statistically variable data ($\vec{\mathbf{y}} \equiv (\mathbf{y}_1, \mathbf{y}_2)$) drawn from three classes.[13] The application domain was human observer performance modeling for acoustical psychophysics experiments. (In prior work, Scurfield investigated a decision rule for three-class classification of univariate data.[17] We will not review that prior work here, because at present we are interested in relating given observer models to the three-class ideal observer model for multivariate observational data, which yield two-dimensional decision variable data by Eq. 3.) In Scurfield's work, no assumptions are made about the decision variables $\mathbf{y}_1$ and $\mathbf{y}_2$; in particular, these decision variables are not assumed to be related in any way to an ideal observer model. This is entirely appropriate given the nature of the problem domain Scurfield investigated — *i.e.*, human observer performance modeling. It can readily be shown, however, that if one chooses to make such assumptions, special cases of the Scurfield model are in fact special cases of an ideal observer decision rule.

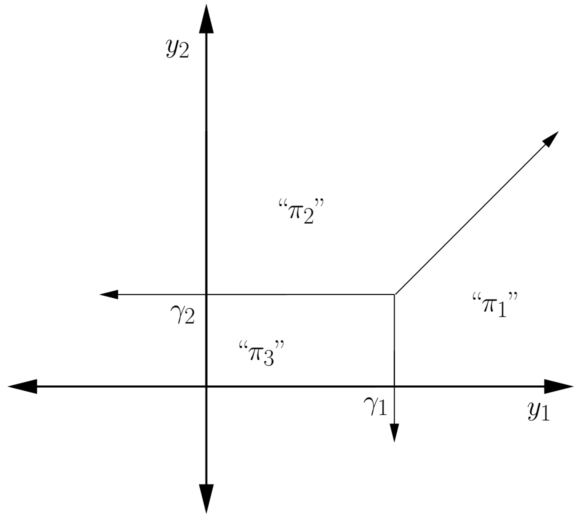

Figure 2. Decision rule investigated by Scurfield, for the decision parameters γ_1 and γ_2.

The Scurfield decision rule is dependent on two decision parameters, which we will call γ_1 and γ_2. The decision rule can be written as

$$\text{decide} \quad d = \pi_1 \quad \text{iff} \quad y_1 - y_2 \geq \gamma_1 - \gamma_2 \quad \text{and} \quad y_1 \geq \gamma_1; \tag{13}$$

$$\text{decide} \quad d = \pi_2 \quad \text{iff} \quad y_1 - y_2 < \gamma_1 - \gamma_2 \quad \text{and} \quad y_2 \geq \gamma_2; \tag{14}$$

$$\text{decide} \quad d = \pi_3 \quad \text{iff} \quad y_1 < \gamma_1 \quad \text{and} \quad y_2 < \gamma_2. \tag{15}$$

This decision rule is illustrated in Fig. 2.

From these relations, one can define the decision boundary lines

$$y_1 - y_2 = \gamma_1 - \gamma_2 \quad \{\text{“1-}\textit{vs.}\text{-2”}\} \tag{16}$$

$$y_1 = \gamma_1 \quad \{\text{“1-}\textit{vs.}\text{-3”}\} \tag{17}$$

$$y_2 = \gamma_2 \quad \{\text{“2-}\textit{vs.}\text{-3”}\}. \tag{18}$$

Note the similarity in form between these equations and Eqs. 7-9. If we choose $\mathbf{y}_1 \equiv \mathrm{LR}_1(\vec{\mathbf{x}})$ and $\mathbf{y}_2 \equiv \mathrm{LR}_2(\vec{\mathbf{x}})$ for some set of observational data $\vec{\mathbf{x}}$, we have a special case of Eqs. 7-9, which is illustrated in Fig. 3.

A second correspondence between Scurfield's decision rule and the ideal observer decision rule can be obtained by taking $\mathbf{y}_1 \equiv \log(\mathrm{LR}_1(\vec{\mathbf{x}}))$ and $\mathbf{y}_2 \equiv \log(\mathrm{LR}_2(\vec{\mathbf{x}}))$; note that a line of the form $\log(\mathrm{LR}_2) = \log(\mathrm{LR}_1) + \alpha$ corresponds to a line of the form $\mathrm{LR}_2 = \beta \mathrm{LR}_1$ for appropriate constants α and β. By inspection, this is again a special case of Eqs. 7-9, which is illustrated in Fig. 4.

Scurfield points out[13] that the observer which maximizes P_C, the "percent correct" or probability of a correct response, is a special case of the ideal observer (*i.e.*, a single operating point achievable by the ideal observer for the given task). This observer follows the Scurfield decision rule model with $\mathbf{y}_1 \equiv \log(\mathrm{LR}_1(\vec{\mathbf{x}}))$ and $\mathbf{y}_2 \equiv \log(\mathrm{LR}_2(\vec{\mathbf{x}}))$, and decision parameters given by $e^{\gamma_1} = P(\pi_3)/P(\pi_1)$ and $e^{\gamma_2} = P(\pi_3)/P(\pi_2)$. It is interesting to note that the Scurfield decision rule model can in fact be used to describe ideal observer performance for an even wider class of operating points, as shown in this section.

4. THE CHAN DECISION RULE

Chan *et al.* are investigating three-class classifiers for computer-aided diagnosis.[14] Their work is motivated by reasoning similar in principle to that which we independently arrived at when we began to consider this problem. In particular, they consider a clinical situation in which observations must be classified as malignant, benign,

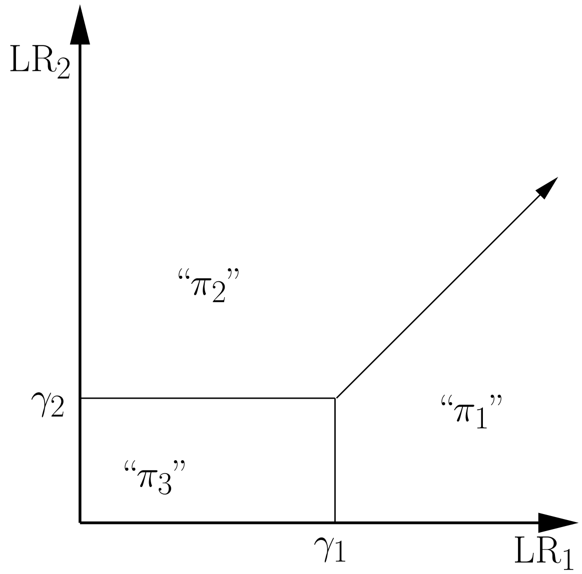

Figure 3. A special case of the ideal observer decision rule, which is a special case of the Scurfield decision rule with $\mathbf{y}_1 \equiv \mathrm{LR}_1(\vec{\mathbf{x}})$ and $\mathbf{y}_2 \equiv \mathrm{LR}_2(\vec{\mathbf{x}})$.

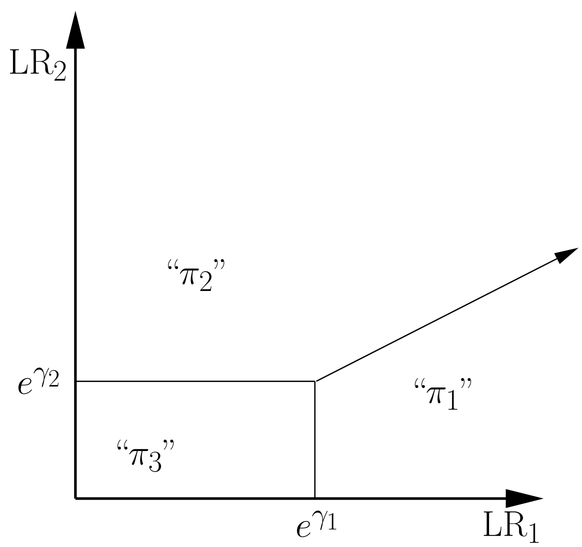

Figure 4. A special case of the ideal observer decision rule which is a special case of the Scurfield decision rule with $\mathbf{y}_1 \equiv \log(\mathrm{LR}_1(\vec{\mathbf{x}}))$ and $\mathbf{y}_2 \equiv \log(\mathrm{LR}_2(\vec{\mathbf{x}}))$.

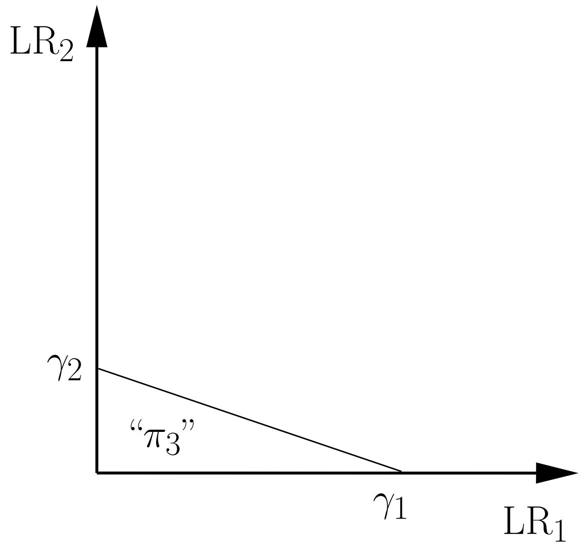

Figure 5. The decision rule investigated by Chan *et al.*, which as they state is a special case of the ideal observer decision rule. Observations in the unlabelled region are decided "not π_3", *i.e.*, either "π_1" or "π_2".

or normal. Because the goal of their work is to optimize the performance of a system to aid a radiologist or clinician, rather than to measure the psychophysical performance of an existing observer, they choose to start explicitly from an ideal observer model in constructing their decision rule.

In order to reduce the complexity of the ideal observer decision rule to manageable proportions, Chan *et al.* impose restrictions on the utilities used by their observer. In their formulation, the class we are labelling π_1 is the benign class; π_2, the normal class; and the malignant class is π_3. They further assume that the possible values of any utility $U_{i|j}$ are restricted to the interval $[0, 1]$. They then set $U_{1|1} = U_{2|2} = U_{3|3} = 1$ (*i.e.*, correctly identifying any case has maximal utility). Furthermore, they require $U_{2|1} = U_{1|2} = 1$ and $U_{1|3} = U_{2|3} = 0$ (*i.e.*, misidentifying a benign case as normal, or vice versa, has no significant cost reducing the utility of such a decision from the maximum, but misclassifying an actually malignant case as benign or normal has the minimum possible utility). Finally, $U_{3|1}$, and $U_{3|2}$ are assumed to have arbitrary values on the open interval $(0, 1)$ (*i.e.*, misclassifying an actually non-malignant case as malignant will have some cost reducing the utility of such a decision from the maximum, but such a misclassification is in some sense "better" than missing an actual malignancy). It is important to note that these assumptions are arguably relevant to a reasonable model of a clinical situation, and are thus of interest beyond their superficial advantage in reducing the degrees of freedom involved in the observer's decision rule. We will, however, only consider the latter issue in the remainder of this section.

Substituting the values of the utilities given above into Eq. 4, we obtain decision boundary lines of the form

$$0\,\mathrm{LR}_1 + 0\,\mathrm{LR}_2 = 0 \qquad \{\text{"1-}vs\text{.-2"}\} \tag{19}$$

$$\frac{(1-U_{3|1})P(\mathbf{t}=\pi_1)}{\alpha}\mathrm{LR}_1 + \frac{(1-U_{3|2})P(\mathbf{t}=\pi_2)}{\alpha}\mathrm{LR}_2 = \frac{P(\mathbf{t}=\pi_3)}{\alpha} \qquad \{\text{"1-}vs\text{.-3"}\} \tag{20}$$

$$\frac{(1-U_{3|1})P(\mathbf{t}=\pi_1)}{\alpha}\mathrm{LR}_1 + \frac{(1-U_{3|2})P(\mathbf{t}=\pi_2)}{\alpha}\mathrm{LR}_2 = \frac{P(\mathbf{t}=\pi_3)}{\alpha} \qquad \{\text{"2-}vs\text{.-3"}\} \tag{21}$$

where $\alpha \equiv 1 + P(\mathbf{t} = \pi_3) - U_{3|1}P(\mathbf{t} = \pi_1) - U_{3|2}P(\mathbf{t} = \pi_2)$. Note that, as Chan *et al.* point out, the "1-*vs.*-2" line is in fact undefined for this choice of utilities, while the "1-*vs.*-3" and "2-*vs.*-3" lines are identical. This is a general consequence of Eqs. 7-9; if any two of these equations yield identical lines, the third line must be either identical to them or undefined. The decision rule considered by Chan *et al.* is illustrated in Fig. 5.

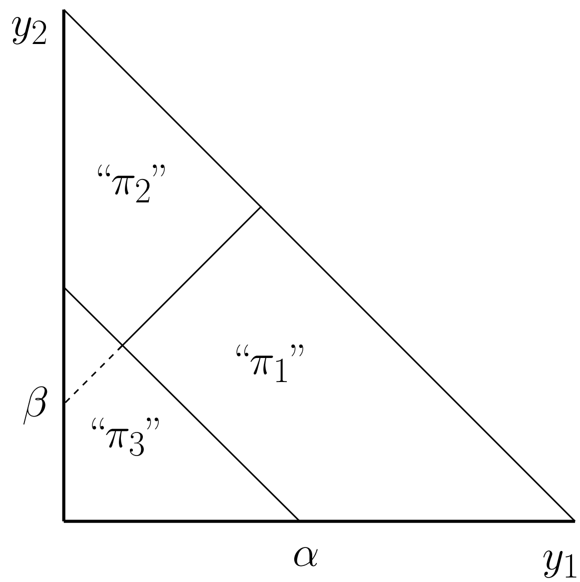

Figure 6. Decision rule investigated by Mossman, for the decision parameters α and β, shown in the *a posteriori* class probability space.

5. THE MOSSMAN DECISION RULE

Mossman investigates a decision rule applied to a set of three decision variables $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$, subject to the constraint

$$\mathbf{y}_1 + \mathbf{y}_2 + \mathbf{y}_3 = 1, \tag{22}$$

as well as $0 \leq \mathbf{y}_i \leq 1 \ \{1 \leq i \leq 3\}$. This is consistent with the constraint on the *a posteriori* class probabilities, $P(\pi_1|\vec{\mathbf{x}}) + P(\pi_2|\vec{\mathbf{x}}) + P(\pi_3|\vec{\mathbf{x}}) = 1$; these quantities are known to be directly related to the likelihood ratio ideal observer decision variables.[18,19] (In this section we will write $P(\pi_i|\vec{x})$ instead of $P(\mathbf{t} = \pi_i|\vec{x})$ for simplicity.) Mossman does not explicitly require, however, that the decision variables in Eq. 22 be the *a posteriori* class probabilities (*e.g.*, they may be noisy estimates of these quantities).

The decision rule considered by Mossman, which depends on two decision parameters α and β, is

$$\begin{array}{llllll} \text{decide} & d = \pi_1 & \text{iff} & y_2 - y_1 \leq \beta & \text{and} & y_3 \leq \alpha; \end{array} \tag{23}$$

$$\begin{array}{llllll} \text{decide} & d = \pi_2 & \text{iff} & y_2 - y_1 > \beta & \text{and} & y_3 \leq \alpha; \end{array} \tag{24}$$

$$\begin{array}{llll} \text{decide} & d = \pi_3 & \text{iff} & y_3 > \alpha. \end{array} \tag{25}$$

where $0 \leq \alpha \leq 1$ and $-1 \leq \beta \leq 1$. From these relations, and given the relation $y_3 = 1 - y_1 - y_2$ from Eq. 22, one can define the decision boundary lines

$$y_1 - y_2 = -\beta \quad \{\text{"1-}vs.\text{-2"}\} \tag{26}$$

$$y_1 + y_2 = 1 - \alpha \quad \{\text{"1-}vs.\text{-3"}\} \tag{27}$$

$$y_1 + y_2 = 1 - \alpha \quad \{\text{"2-}vs.\text{-3"}\}. \tag{28}$$

This decision rule is illustrated in Fig. 6. Note that, similar to the Chan *et al.* decision rule, the "1-*vs.*-3" and "2-*vs.*-3" decision boundary lines are identical.

We now consider a special case of the Mossman decision rule in which $\mathbf{y}_1 = P(\pi_1|\vec{\mathbf{x}})$, $\mathbf{y}_2 = P(\pi_2|\vec{\mathbf{x}})$, and $\mathbf{y}_3 = P(\pi_3|\vec{\mathbf{x}})$ for some observational data vector $\vec{\mathbf{x}}$. This version of the decision rule is illustrated in Fig. 7.

Although the Mossman decision rule appears similar in form to the ideal observer decision rule, recall from Sec. 4 that if two of the decision boundary line equations are identical, the third must yield a line identical to

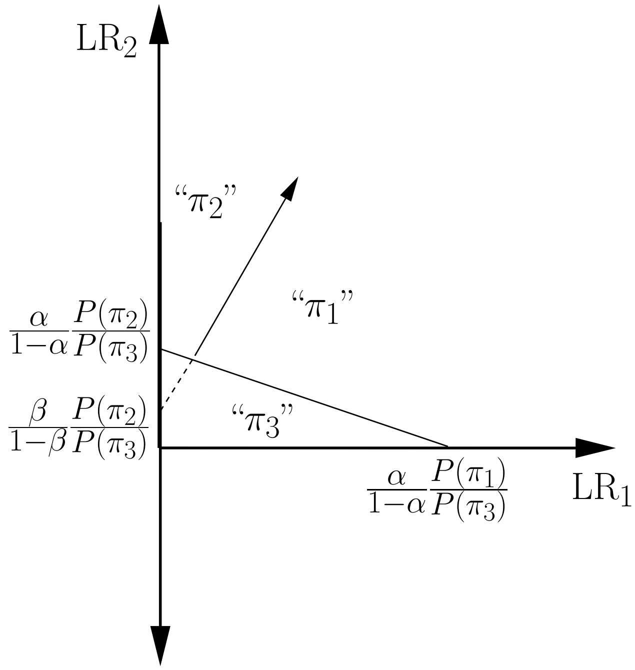

Figure 7. Decision rule investigated by Mossman, for the decision parameters α and β, shown in likelihood ratio space.

the first two or be undefined. Another way to see this is to note that the coefficients of Eq. 9 are differences of the corresponding coefficients of Eqs. 7 and 8. If the coefficients of Eqs. 8 and 9 are identical, it must be the case that the coefficients of Eq. 7 are all zero. For the Mossman decision rule, this would require $1+\beta=0$, $1-\beta=0$, and $\beta=0$ simultaneously, which is clearly impossible. It follows that the decision rule considered by Mossman cannot represent possible ideal observer performance for any choice of the utilities $U_{i|j}$ in Eqs. 1 and 2.

6. DISCUSSION AND CONCLUSIONS

We examined three decision rules proposed recently for three-class classification tasks by different researchers. The basis for our evaluation was ideal observer decision theory, primarily because our own interest in the three-class classification task is its possible application to CAD.

Although this is not the most general approach to three-class classification, the three-class classification task is difficult enough that it is perhaps worth making any attempt to analyze, from a single point of view, the work of the relatively few researchers investigating this problem.

In particular, Scurfield points out[13] that his proposed decision rule is in fact an ideal observer decision rule for a single ideal observer operating point, namely the observer which maximizes the probability of any correct response (or "percent correct" or P_C). We were able to show that, under various assumptions, a larger set of such correspondences between the Scurfield observer and the ideal observer exists.

Chan *et al.* are working on the application of three-class classification to CAD, and thus explicitly take the ideal observer as the starting point in the development of their decision rule.[14] Although this rendered our analysis of that decision rule in terms of ideal observer decision theory largely trivial, it provided an intuitive basis for understanding the results of similar analysis of the Mossman decision rule, namely the conclusion that the latter does not correspond to ideal observer behavior for any possible values of the utilities used by the ideal observer. However, we note that the structure of the Mossman decision rule — a simple sequence of thresholds on single decision variables — may indeed serve as a reasonable model for human observer performance in certain situations, *e.g.*, differential diagnosis.

ACKNOWLEDGMENTS

The authors thank Vit Drga for bringing Brian Scurfield's work to their attention, and thank both Vit Drga and Brian Scurfield for helpful conversations concerning that work. The authors thank Heang-Ping Chan and Berkman Sahiner for helpful conversations concerning their work.

This work was supported in parts by grant W81XWH-04-1-0495 from the US Army Medical Research and Materiel Command (D. C. Edwards, principal investigator). Charles E. Metz is a shareholder in R2 Technology, Inc. (Sunnyvale, CA).

REFERENCES

1. U. Bick, M. L. Giger, R. A. Schmidt, R. M. Nishikawa, D. E. Wolverton, and K. Doi, "Automated segmentation of digitized mammograms," *Acad. Radiol.* **2**, pp. 1–9, 1995.
2. F.-F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," *Med. Phys.* **18**, pp. 955–963, 1991.
3. F.-F. Yin, M. L. Giger, C. J. Vyborny, K. Doi, and R. A. Schmidt, "Comparison of bilateral-subtraction and single-image processing techniques in the computerized detection of mammographic masses," *Invest. Radiol.* **28**, pp. 473–481, 1993.
4. F.-F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique," *Med. Phys.* **21**, pp. 445–452, 1994.
5. M. A. Kupinski, *Computerized Pattern Classification in Medical Imaging.* Ph.D. thesis, The University of Chicago, Chicago, IL, 2000.
6. Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," *Acad. Radiol.* **5**, pp. 155–168, 1998.
7. Z. Huo, M. L. Giger, and C. E. Metz, "Effect of dominant features on neural network performance in the classification of mammographic lesions," *Phys. Med. Biol.* **44**, pp. 2579–2595, 1999.
8. Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, and C. E. Metz, "Computerized classification of benign and malignant masses on digitized mammograms: A study of robustness," *Acad. Radiol.* **7**, pp. 1077–1084, 2000.
9. Z. Huo, M. L. Giger, and C. J. Vyborny, "Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis," *IEEE Trans. Med. Imag.* **20**, pp. 1285–1292, 2001.
10. Z. Huo, M. L. Giger, C. J. Vyborny, and C. E. Metz, "Breast cancer: Effectiveness of computer-aided diagnosis — Observer study with independent database of mammograms," *Radiology* **224**, pp. 560–568, 2002.
11. H. L. Van Trees, *Detection, Estimation and Modulation Theory: Part I*, John Wiley & Sons, New York, 1968.
12. C. E. Metz and X. Pan, "'Proper' binormal ROC curves: Theory and maximum-likelihood estimation," *J. Math. Psychol.* **43**, pp. 1–33, 1999.
13. B. K. Scurfield, "Generalization of the theory of signal detectability to n-event m-dimensional forced-choice tasks," *J. Math Psychol.* **42**, pp. 5–31, 1998.
14. H.-P. Chan, B. Sahiner, L. M. Hadjiiski, N. Petrick, and C. Zhou, "Design of three-class classifiers in computer-aided diagnosis: Monte carlo simulation study," in Proc. SPIE Vol. 5032 *Medical Imaging 2003: Image Processing*, Milan Sonka and J. Michael Fitzpatrick, eds., pp. 567–578, (SPIE, Bellingham, WA), 2003.
15. D. Mossman, "Three-way ROCs," *Med. Decis. Making* **19**, pp. 78–89, 1999.
16. D. C. Edwards, C. E. Metz, and M. A. Kupinski, "Ideal observers and optimal ROC hypersurfaces in N-class classification," *IEEE Trans. Med. Imag.* **23**, pp. 891–895, 2004.

17. B. K. Scurfield, “Multiple-event forced-choice tasks in the theory of signal detectability,” *J. Math Psychol.* **40**, pp. 253–269, 1996.
18. M. A. Kupinski, D. C. Edwards, M. L. Giger, and C. E. Metz, “Ideal observer approximation using Bayesian classification neural networks,” *IEEE Trans. Med. Imag.* **20**, pp. 886–899, 2001.
19. D. C. Edwards, L. Lan, C. E. Metz, M. L. Giger, and R. M. Nishikawa, “Estimating three-class ideal observer decision variables for computerized detection and classification of mammographic mass lesions,” *Med. Phys.* **31**, pp. 81–90, 2004.