XXI IMEKO World Congress "Measurement in Research and Industry"
August 30 – September 4, 2015, Prague, Czech Republic

ADAPTIVE CELL DETECTION BY APPLICATION OF ROC METHODOLOGY

Giampaolo E. D'Errico

Istituto Nazionale di Ricerca Metrologica (INRIM), Torino (I), e-mail: g.derrico@inrim.it

Abstract – Receiver operating characteristic (ROC) methodology is finalized to modelling and implementation of an adaptive detection process with application to cellular micrographs. Numerical examples are elaborated to illustrate how the performance of a cell recognition task can be modulated according as a 4-point rating scale of process dependability.

Keywords: cell micrographs, ROC methodology, adaptive image thresholding

1. INTRODUCTION

In cellular biology experiments, proliferation and growth of cultured cells are consequential observables to monitor and analyse, so that timely decisions can be taken during the culture process. On one hand, a synthetic index for tracking the experiment's performance is the confluence of cells, a quantity expressing the relative occupancy of cells over a surface (typically a dish). On the other hand, if precise enumeration of cells is the needed quantity, an exhaustive recognition of entities one by one is required over the microscopy image being inspected.

This research work is addressed to the segmentation of 2D images composed of regions showing diverse grey levels by image thresholding techniques. In particular, images of interest here are cell micrographs. Segmentation is partitioning the whole collection of regions into two sub-collections, i.e. to obtain a binary image where foreground entities (objects of interest) and background entities are complementary image's content. A review of image segmentation techniques can be found in [1].

With application to cellular images, the performance of a segmentation process is therefore affected by the threshold adopted to decide whether to assign a detected object to the partition of cells or not. SW procedures are developed in the bio-informatics area to automatize the process [2], and the ability of algorithms is being compared to skilled human operators (e.g., with application to fluorescence micrographs, [3]).

In the present paper, the confidence attributed to the decision-making is modelled in terms of the receiver operating characteristic (ROC) methodology. A tutorial on ROC can be found in [4]. While a ROC-based approach with applicability to cell confluence estimation to cell micrographs was finalized by the author in [5], in the following a numerical study is developed and some examples are present and discussed. In ROC terms, the rational of this study is that adaptive thresholding can be modulated over a performance rating scale, eventually mapping the comparative dependability of segmentation processes. The numerical examples are processed by use of a SW made available on the internet by the Johns Hopkins University School of Medicine (Baltimore, Maryland, USA), namely the JROCFIT calculator [6].

The next section gives a brief account of the approach to cellular images segmentation developed on the base of the ROC methodology; in view of the performance evaluation of the segmentation process, the role of pre-process information about images content is remarked; the notion of the area under the ROC curve (AUC) and its interpretation is also introduced.

Using that theoretical tool, in Section 3 a numerical treatment is presented and discussed with application to a cell detection task, where—in the absence of a "gold standard"—each inspected entity is adaptively categorized in terms of four qualitative labels, namely: definitely not a cell; probably not a cell; probably a cell; definitely a cell.

2. ROC CURVE AND AREA UNDER THE CURVE

Segmentation is based on the following facts. In images under treatment, the grey levels of points belonging to the object of interest are sensibly different from the grey level pertaining to the background. Such an image can be characterized in terms of a 2-place function $g = g(x, y)$, where (x, y) is a picture element in spatial coordinates (a pixel) and g represents the grey level associated to (x, y). A grey value g_{th} is chosen as a threshold to discriminate objects from other (uninteresting) pixel regions (see [7] for more technical details).

Several approaches have been developed to improve the choice of a threshold [1], including those based on heuristic criteria (e.g., minimization of measure of fuzziness [8]) or decision-theoretical criteria (e.g., the minimum error criterion [9]). In fact, the choice of a value for g_{th} affects the (expected) rates of correct or erroneous assignments; if g_{th} is changed, these rates change consequently. A model to describe such changes is the ROC curve.

To illustrate the model based on the receiver operating characteristic, the following variables are introduced: *TP*, true positives (the region belongs, and is correctly assigned, to the object); *TN*, true negatives (the region belongs, and is correctly assigned, to the image background); *FP*, false positives (the region belongs to background, but is erroneously assigned to the object); and *FN*, false negatives

(the region belongs to the object, but its recognition is missed). Moreover, if the total of positive cases is denoted by $Tot^+ = TP + FN$ and total of negative cases by $Tot^- = TN + FP$, two nominal characteristics of the recognition process can be defined in terms of above introduced rates, namely: the sensitivity, $Se = TP/Tot^+$; and the specificity, $Sp = TN/Tot^-$.

In fact, sensitivity and specificity are related to each other—as one increases the other decreases—so that for each threshold value g_{th}, *Se* and *Sp* values are coupled together. Plotting the variation with g_{th} of the couples $(1 - Sp, Se)$ on a Cartesian graph gives rise to a ROC curve. Axes in a ROC graph are customary labelled false positives rate, $FPR = 1 - Sp$ (the abscissa), and true positives rate, $TPR = Se$ (the ordinate). The curve rises from the point (0, 0) (no false positive cases are only obtained at the cost of no true positives cases) and reaches the point (1, 1) (100% of true positive rate is obtained only at the cost of 100% false positive rate) with the varying slope $Se/(1 - Sp)$, that is the likelihood ratio $LR = TP(TP + FN)^{-1}(TN + FP)FP^{-1}$.

However, these rates represent nominal performance indices, unable to express and to predict the actual performance of the process—that depends also on the characteristic of the image being processed. To estimate such a performance, two other variables are needed—namely the positive predictive value (*PPV*) and the negative predictive value (*NPV*)—, that in terms of probabilities are introduced as follows.

Let c represent a binary variable such that $c = 1$ if the image region is a cell, otherwise $c = 0$; let r represent a bivalent—yes (Y)/no (N)—variable such that $r = Y$ if the response of visual inspection is positive (for the region to be classified a cell), otherwise $r = N$. Let $\Pi(c = 1)$ and $\Pi(c = 0) = 1 - \Pi(c = 1)$ denote the pre-process (prior) probability of a region to be classified as a cell or as a non-cell, respectively. Using this notation, *Se* and *Sp* can be reformulated in terms of conditional probability:

$$Se = \Pi(r = Y|c = 1) \quad (1)$$

$$Sp = \Pi(r = N|c = 0) \quad (2)$$

Based on Eq. (1) and Eq. (2), *PPV* and *NPV* can thus be defined in terms of inverse probability (Bayes rule):

$$PPV = \Pi(c = 1|r = Y) = Se\Pi(c = 1)/\Pi(r = Y) \quad (3)$$

$$NPV = \Pi(c = 0|r = N) = Sp\Pi(c = 0)/\Pi(r = N) \quad (4)$$

where:

$$\Pi(r = Y) = Se\Pi(c = 1) + (1 - Sp)\Pi(c = 0) \quad (5)$$

$$\Pi(r = N) = 1 - \Pi(r = Y) = Sp\Pi(c = 0) + (1 - Se)\Pi(c = 1) \quad (6)$$

With application to cell confluence estimation, the prior probabilities $\Pi(c = 1)$ and $\Pi(c = 0) = 1 - \Pi(c = 1)$ can be used to translate numerically the pre-process information about the confluence from, e.g., qualitative expressions such as “low”, “medium”, or “high” confluence.

In the ROC framework, the accuracy of a state discrimination process is modulated in terms of sensitivity and specificity pairs as plotted in the ROC curve. Moreover, the area under the curve (AUC) has an interesting probabilistic interpretation that helps clarifying the process of thresholding.

Following the treatment presented in [10], let s_1, s_0 be two continuous random variables introduced to denote the strength of perception leading to assignment of an image region to cell population or to background, respectively. Modelling the observer’s criterion in term of a threshold s_{th} the probabilities of assignment translate into: $\Pi(r = Y|c = 1) = \Pi(s_1 \geq s_{th})$, $\Pi(r = Y|c = 0) = \Pi(s_0 \geq s_{th})$. The area under a ROC curve plotted on axes $x = \Pi(r = Y|c = 1)$, $y = \Pi(r = Y|c = 0)$ is therefore identical to the area under a curve plotted on axes $x = \Pi(s_1 \geq s_{th})$, $y = \Pi(s_0 \geq s_{th})$. Provided s_1 and s_0 are mutually statistically independent, this area is:

$$\mathrm{AUC}(s_0, s_1) = \Pi(s_0 < s_1) \quad (7)$$

This model is also known as a probabilistic “stress-strength” model, where s_0 and s_1 (in this model representing stress and strength, respectively) have a joint distribution $\pi(s_0, s_1) = f(s_0)g(s_1)$. Putting $G_{s_1}(s_0) = \int_{-\infty}^{s_0} g(s_1)d\,s_1$:

$$\Pi(s_0 < s_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{s_0} \pi(s_0, s_1)ds_0ds_1 = \int_{-\infty}^{\infty} G_{s_1}(s_0)f(s_0)d\,s_0 \quad (8)$$

Note that there is no assumption about the form of distributions of s_0, s_1: it can be pointed out that if they are identically distributed, then $\mathrm{AUC}(s_0, s_1) = 0.5$, meaning the detection is in fact a random process and its result is uninformative. Moreover, when $\mathrm{AUC}(s_0, s_1) = 1$, s_0 is distributed everywhere below s_1, meaning the observer is able to perform the task ideally in an errorless way; finally, the larger the AUC, the better the performance: the ROC curves can be rated accordingly.

The meaning of the area under the curve can be clarified in the framework of the so-called two alternative forced choice (2AFC) technique. Conceptually, from the 2AFC point of view the detection process can be thought of as if the observer be presented an ordered couple of regions to be matched, one with a prototypical representative of cells, the other with a prototypical background region [11]. The observer decision is to assign a couple of degrees (on a mutually independent rating scale in the real interval [0, 1]) of similarity between a region and its prototypical image. These degrees are probabilities in the stress-strength model; they can also be interpreted in terms of membership degrees in the framework of a fuzzy-set based model.

3. APPLICATION OF ROC-BASED METHODOLOGY TO CELL DETECTION

3.1. Numerical examples

As to fluorescence micrographs, sources of possible misinterpretation include problems with sample staining, inconstancy of cell shapes, non uniformity illumination, and out of focus blurring (see, e.g., [3]): if a sort of gold standard (“ground truth”) is referred to, it is the outcome of a manual segmentation obtained by the naked eye. However, the gold standard—being obtained by inspection

accomplished by human experts—may vary according to the expert's operating characteristic. Nevertheless, related variations can be accounted for by application of ROC-based methodology.

Focusing on ROC analysis, a data format of interest for application to cell recognition is represented by cumulative *TPR* and *FPR* categorized according to an ordinal rating scale. This format is supported by the JROCFIT web-based calculator [6], that is also able to fit a ROC curve in the absence of a "gold standard" or "ground truth"—the idea is to assign Tot^+ and Tot^- arbitrarily large numbers (see format 4 in [6] for details) when the "true" total number of positive and negative cases is unknown. Advancing in the direction of this clue, the integration of pre-process information about cell confluence can be implemented by use of the JROCFIT calculator.

The idea of supplying to the lack of a ground truth can be illustrated by an example, where pre-process estimations of "low", "medium", and "large" cell confluence are expressed in terms of the ratio $Tot^+/(Tot^+ + Tot^-)$. This ratio, taken over an arbitrary large number of $Tot^+ + Tot^-$ gives also an estimation of the prior probability of an area being a cell inside the whole image under inspection: this is the probability $\Pi(c = 1)$ affecting quantities expressed by Eqs. (1–4), in particular *Se* and *Sp* used to plot ROC curves.

The probabilities for an inspected entity to be classified as cell or as a non-cell can be mapped into cut-off points that, in turn, can be referred to for rating the variable performance attainable by the operator in terms of dependability levels. For example, a ROC-based graph can report results of a 4-point rating process according to four qualitative categories (say category #0, #1, #2, and #3) ordered in increasing order of dependability.

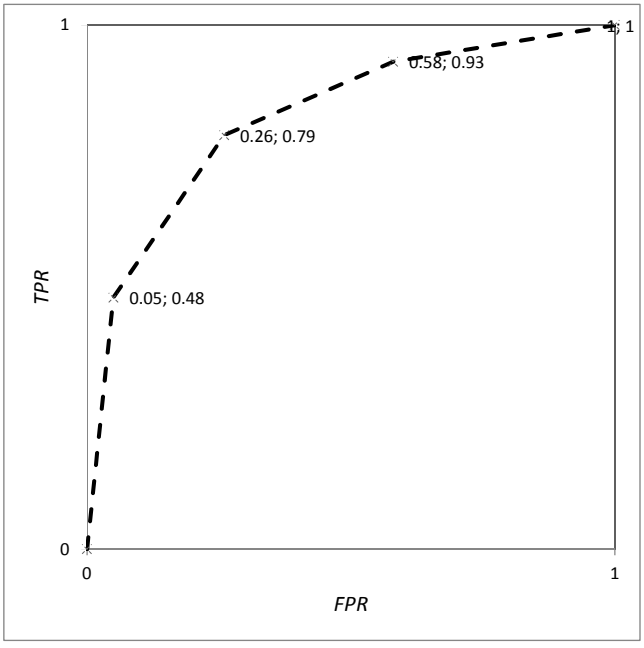

Fig. 1. Roc plot of a 4-level rating scale, with application to cell vs non-cell recognition.

With application to cell detection, each inspected entity is categorized on the base of the following criterion: #0: definitely not a cell; #1: probably not a cell; #2: probably a cell; #3: definitely a cell—the assignment criterion being the result of a combination of rules implemented to cope with sources of misinterpretation of optical observations.

Fig. 1 shows a ROC graph relevant to a cell recognition process. Coordinates relevant to said 4-point rating—listed in order of increasing dependability level—are (1.0, 1.0), (0.58, 0.93), (0.26, 0.78), (0.05, 0.48): abscissa and ordinate are expressed in terms of cumulative values of *FPR* and *TPR*, respectively. This list represents the percentage of all supposedly "true" non-cells (abscissa) and cells (ordinate), that were assigned to category #*n* (n = 0, 1, 2, 3), i.e. with a dependability rating at least as high as the *n*th dependability level.

The exact number of "true" cells and "true" non-cells is in fact unknown; however cell confluence, initially classified by use of qualitative labels, namely "low", "medium" or "high" confluence, can be estimated—according to format 4 in JROCFIT calculator—by use of arbitrary large numbers.

These ("high" and "low") labels are numerically translated in the following terms: given a total of $Tot^+ + Tot^-$ = 5000 cases, "low" confluence is a situation where 20% of cases are estimated actually cells; "high" confluence is the opposite situation: 20% of cases are estimated actually non-cells.

If by convention it is assumed that a cell is a positive case, let *TP* (*FP*, respectively) stand for the number of cells correctly (erroneously, respectively) recognized, and let *TN* (*FN*, respectively) stand for the number of non-cells correctly (erroneously, respectively) recognized; $Tot^+ = TP + FN$ and $Tot^- = TN + FP$.

3.1. Results and discussion

Table 1. 4-category classification, using JROCFIT [6]: the higher the category number, the higher the dependability of the response.

	"low" (20%) confluence		"high" (80%) confluence	
category	non-cells *TN+FP*	cells *TP+FN*	non-cells *TN+FP*	cells *TP+FN*
#0	1680	70	420	280
#1	1280	150	320	600
#2	840	300	210	1200
#3	200	480	50	1920
	Tot^- = 4000	Tot^+ = 1000	Tot^- = 1000	Tot^+ = 4000

Response data of JROCFIT calculator [6], representing results of the cell recognition exercise with the characteristics described above, are as shown in Tab. 1 and in Fig. 2.

Summary statistics and global performance indices are as follows:

- "low" confluence situation: TP = 780, FN = 220 (cells missed), TN = 2960, FP = 1040 (non-cells missed), $TP + TN$ = 3740 (cases correctly recognized), Se = 78%, Sp = 74%, $(TP + TN)/(Tot^+ + Tot^-)$ = 74.8% (accuracy);

· "high" confluence situation: TP = 3120, FN = 880, TN = 740, FP = 260, $TP + TN$ = 3860, Se = 78%, Sp = 74%, $(TP + TN)/(Tot^{+} + Tot^{-})$ = 77.2%.

The global specificity (TN/Tot^{-}) and the global sensitivity (TP/Tot^{+}) are characteristics of the recognition process that are not affected by initial estimate of confluence, that instead influences the accuracy.

The curves in top panel and bottom panel of Fig. 2 are obtained with reference to "high" and "low" confluence, respectively, with application to the cell recognition process described in Tab. 1 and Fig. 1.

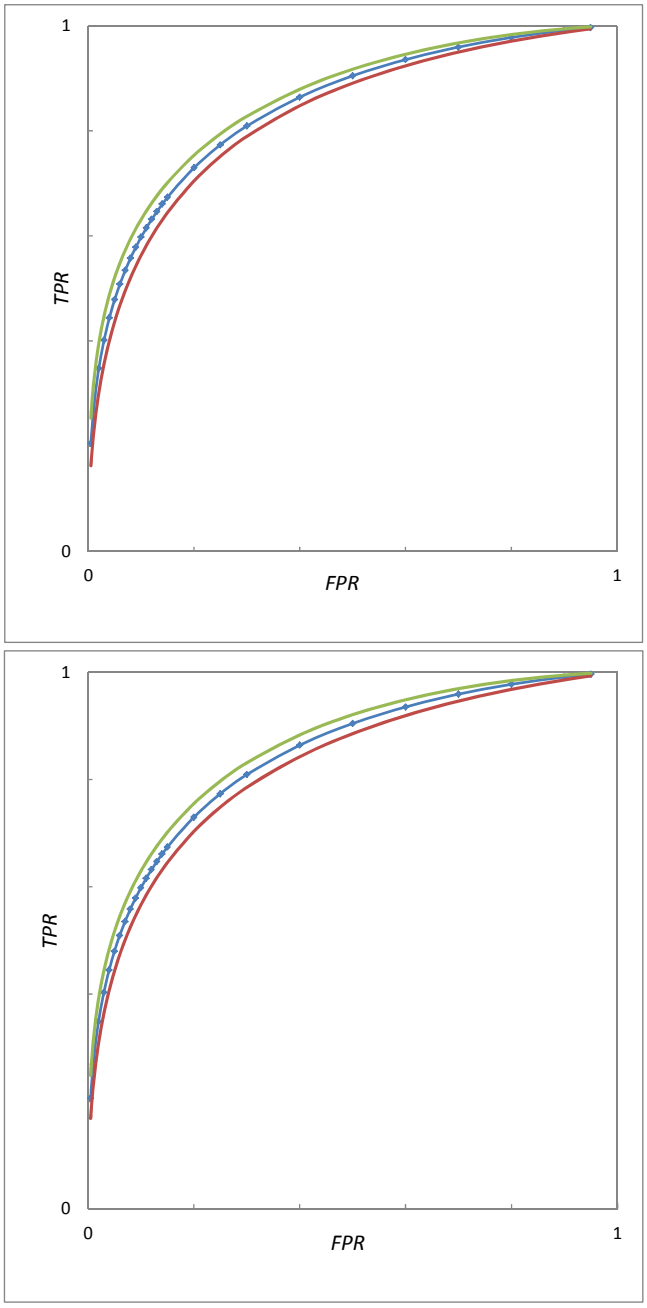

Fig. 2. Roc curves fitted to data plotted in Fig. 1, showing lower and upper bounds (95% confidence interval) for "high" confluence (top panel) and "low" confluence (bottom panel) situations.

The area under the empiric ROC curve, AUC_e = 0.823 and the area under the fitted ROC curve, AUC_f = 0.8431 are the same for both confluence situations, however the standard deviation of AUC_f is 0.0077 for the "low" confluence and 0.0067 for the "high" confluence (AUC is a global descriptor of the modelled performance: AUC = 1 represents ideally errorless performance).

From Fig. 2 it can be also observed that the lower and upper bounds (95% confidence interval) depend on the estimated confluence: comparing "low" and "high" situations, bounds result closer to the fitted curve relevant to the "low" confluence data. This observation is confirmed in terms of the accuracy calculated as the ratio:

$$\alpha = (TP + TN)/(Tot^{+} + Tot^{-}) \qquad (8)$$

If accuracy is a criterion to asses the performance of the cell recognition process, it can be remarked that the performance obtained in both situations could be further improved—provided a more refined rating process is still possible: however this provision is the core of the problem. In fact, given a pre-process estimation of distributions of Tot^{+} and Tot^{-} (here summarized in terms of the descriptors "low" and "high" confluence), the probability of TP and TN, is related to the attitude of the observer, who is discriminating cells vs non-cells in the lack of a "ground truth". A sort of gold standard (presumptive "ground truth") is just an acknowledged result of visual image segmentation attainable after delineation of cell boundaries by the naked eye. Assuming one nucleus per cell for the biotype under observation, cells affected by ambiguous or undistinguishable boundaries and those showing multiple nuclei are assigned to intermediate categories (in this examples, category #1: probably not a cell, and category #2: probably a cell).

From the point of view of the overall process dependability, it is worth of remarking that intermediate assignments make the decision-making process adaptive to an extent that is constrained by the observer's skill represented by the ROC curve (outcomes of human decision-making may vary according to subjective criteria too).

In the above mentioned 2AFC framework, prototypical representatives of cells and of non-cells are referred to in the implementation of a dedicated decision-making task. However, such a prototype (a surrogate for absent gold standard) is just a receiver (i.e., the effective observer) operating characteristic. The ROC-based analysis may help individuating weak points in the categorization design, in order to ameliorate criteria and rules to be implemented in the decision-making procedure. Accordingly, the ROC-based approach has been here managed as a tool to provide a flexible model of thresholding criteria for image segmentation implemented by human operators, in the absence of a "gold standard".

4. CONCLUSION

An approach to modelling and implementation of a segmentation process with application to cellular images has been developed on the base of the ROC methodology.

The relevance of pre-process information for performance evaluation has been pointed out and integrated in a probabilistic model in terms of baseline estimates.

A four-point rating process has been modelled in terms of cut-off points over ROC curves.

A numerical treatment has been implemented focusing on a cell detection task, where—in the absence of a "gold standard"—each inspected entity is adaptively categorized in terms of four qualitative labels, namely: "definitely not a cell", "probably not a cell", "probably a cell", and "definitely a cell".

Some numerical examples have been elaborated by use of the JROCFIT web-based calculator [6], taking into account two pre-process scenarios corresponding to initial cell confluence estimations, labelled "low" (20%) and "high" (80%), respectively.

Relevant results have been illustrated and discussed. From the point of view of this specific application, it can be concluded that the ROC tool can be used to provide an evaluation—in terms of global specificity, sensitivity, accuracy, AUC—of the process performance obtained in both "low" and "high" confluence scenarios.

From a more general point of view, it is worth of remarking that flexible categorization in terms of intermediate assignments (e.g., by categories "probably not a cell" and "probably a cell") allows for adaptive decision-making, so that the observer's skill is fully exploited according to the peculiar ROC curve.

REFERENCES

[1] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and qualitative performance evaluation", *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–165, 2004.

[2] K. W. Eliceiri et al., "Biological imaging software tools", *Nat Methods*, vol. 9, no. 7, pp. 697–710, 2013.

[3] T. W. Nattkemper, T. Twellmann, H. Ritter and W. Schubert, "Human vs. machine: evaluation of fluorescence micrographs", *Computers in Biology and Medicine*, vol. 33, pp. 21–43, 2003.

[4] C. E. Metz, "ROC analysis in medical imaging: a tutorial review of the literature", *Radiol Phys Technol*, vol. 1, pp. 2–12, 2008.

[5] G. E. D'Errico, "Receiver operating characteristic: a tool for cell confluence estimation", *Proceedings IEEE International Symposium on Medical Measurements and Applications (MEMEA 2015)*, pp. 576–579, Torino (I), May 7-9, 2015.

[6] J. Eng, "ROC analysis: web-based calculator for ROC curves", http://www.jrocfit.org (retrieved 4 Dec 2014).

[7] N. R. Pal and S. K. Pal, "A review on image segmentation techniques", *Pattern Recognition*, vol. 26, no. 9, pp. 1277–1294, 1993.

[8] L.-K. Huang and M.-J. J. Wang, "Image thresholding by minimizing the measures of fuzziness", *Pattern Recognition*, vol. 28, no. 1, pp. 41–51, 1995.

[9] J. Kittler and J. Illingworth, "Minimum error thresholding", *pattern recognition*, vol. 19, no. 1, pp. 412–417, 1986.

[10] D. Bamber, "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph", *Journal of Mathematical Psychology*, vol. 12, pp. 387–415, 1975.

[11] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve", *Radiology,* vol. 1, pp. 29–36, 1982.