34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

A NEW ROC ANALYSIS METHOD CONSIDERING THE CORRELATION BETWEEN NEIGHBORING PIXELS

Xin Liu and Imam Samil Yetik

Medical Imaging Research Center, Illinois Institute of Technology, Chicago, IL, USA.

ABSTRACT

In this paper, we introduce a novel receiver operating characteristic (ROC) analysis method that considers spatial correlation between pixels to evaluate classification algorithms. ROC analysis is one of the most important tools in the evaluation of medical images and computer aided diagnosis (CAD) systems. It provides a comprehensive description of the detection accuracy of the test image. To evaluate the localization performance, operating points of ROC curves are obtained based on the classification results of individual pixels. To this date, the confidence level or intensity value of each pixel is assumed to be independent within the image. However, this assumption is not satisfied in real problems. In this paper, a new ROC analysis algorithm that considers the correlation between neighboring pixels is proposed. Our results show that the new ROC curves provide a more accurate evaluation of the test image.

Index Terms— image evaluation, ROC analysis, spatial correlation.

1. INTRODUCTION

Receiver operating characteristic (ROC) analysis is one of the most important tools to assess medical images and classification algorithms used for e.g. tumor localization. It comes from the fundamental principles of statistical decision theory [1] and signal detection theory [2]. ROC analysis evaluates the classification performance in terms of the ability to use the image data to classify the pixels as “object” or “background” with respect to a particular classification task. It can be used to assess the performance in binary classification problems. The ROC curve displays the relationship between true positive fraction (TPF) and false positive fraction (FPF). TPF is the fraction of actually object pixels correctly classified as “object”, and FPF is the fraction of actually background pixels incorrectly classified as “object”. TPF is also called sensitivity, whereas FPF is equivalent to 1-specificity. An ROC curve demonstrates the tradeoff between sensitivity and specificity that a classification algorithm allows as the discrimination threshold of classified object and classified background varies.

There are two alternative methods commonly used to generate ROC curves: i) a nonparametric approach that the operating points obtained by applying a successively threshold (the value above which a result is classified as positive) to the data are plotted, ii) a parametric approach which is based on the model of signal detection theory, assuming the classification results, or a monotonic transformation of them belong belong to certain distributions, and produces smooth ROC curves. It is widely accepted that the second method provides reliable interpolation between the empirical combinations of TPF and FPF that calculated directly from the test image. To fit a smooth ROC curve, maximum likelihood estimation (MLE) and binormal model are commonly used [3]. All of these models mentioned above assume that the test data (intensity value) is uncorrelated.

Historically, all these ROC analysis were developed to evaluate the detection performance of human observers and computer algorithms without considering the possible position and size of the detection task in each image. To further evaluate the localization performance, location receiver operating characteristic (LROC) [4] and free-response operating characteristic (FROC) were developed [5]. LROC and FROC analysis is an active research topic and has been applied to evaluate many CAD systems, especially for mammography [6, 7]. FROC and LROC provide potentially greater statistical power than conventional ROC analysis, but they depend on the amount of location error that is allowed by the data analyst. These methods are also strongly dependent on the similarity measures and various parameters of these measures.

In order to conquer these disadvantages, ROC analysis is performed based on each pixel instead of each image to evaluate the location performance of computer algorithms [8]. That is, the ROC analysis is applied to each pixel to evaluate a particular imaging modality or CAD algorithm. In this way, the ROC analysis evaluate both the detection and localization performance of the image or algorithm. All the available techniques for ROC curve fitting could be employed. Unfortunately, for those curve fitting methods, the pixels are assumed to be spatially independent. However, this assumption may not be satisfied in real problems. In this paper, we propose to consider the spatial correlation between neighboring pixels and create new ROC curves providing a more accurate evaluation of the test image or the localization algorithm.

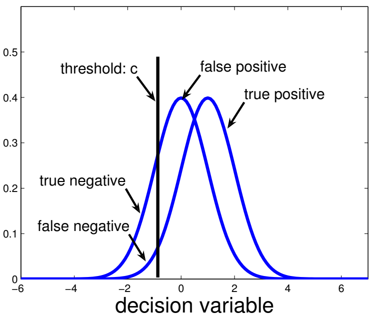


Fig. 1. Illustration of the model used in ROC curve fitting.

The outline of the paper is as follows. Section 2 describes the conventional ROC curve, MLE and binormal model for curve fitting. Section 3 describes the proposed ROC method based on a conditional probabilistic model. In Section 4, the conventional and the new ROC model are applied to test images to show that proposed methods can better quantify the classification power of the test image. Conclusions and discussions are provided in Section 5.

2. THE CONVENTIONAL ROC CURVE

In this section, the most commonly used method: MLE and binormal model to generate conventional ROC curves is described. In order to be consistent with other ROC analysis in the literature, we call the classification result a decision variable, and consider it to be continuously-distributed data. The simplest approach to fit the empirical ROC curve is to estimate the means and standard deviations of the actually negative and actually positive decision variable distributions directly from the test result data, and calculate the ROC curve by assuming the form of the decision variable distributions.

Suppose the decision variable (response or intensity value) is x, the conditional distribution functions of x for actually background pixels and actually object pixels are $p(x|x \in S_1)$ and $p(x|x \in S_2)$. In order to obtain a smooth ROC curve to fit the test data, a mathematical form of the two decision variable distributions should be assumed. Although many assumptions are possible, the most widely used assumption is the "binormal" model which has been found empirically to provide satisfactory ROC curves in a variety of situations. Fig. 1 is an illustration of this model. To plot a ROC curve, we calculate the FPF and TPF corresponding to each setting of the threshold c as follows:

$$\begin{aligned} \text{FPF}(c) &= \text{Prob}(x > c|x \in S_1) = 1 - \text{Prob}(x \le c|x \in S_1) \\ &= 1 - \int_{-\infty}^{c} p(x|x \in S_1)dx, \\ \text{TPF}(c) &= \text{Prob}(x > c|x \in S_2) = 1 - \text{Prob}(x \le c|x \in S_2) \\ &= 1 - \int_{-\infty}^{c} p(x|x \in S_2)dx. \qquad (1) \end{aligned}$$

Assuming $p(x|x \in S_1)$ and $p(x|x \in S_2)$ are normal distributions with means and variance: μ_1, σ_1^2 and μ_2, σ_2^2. Then, we have:

$$\begin{aligned} \text{FPF}(c) &= 1 - \Phi(\frac{c - \mu_1}{\sigma_1}) = \Phi(\frac{\mu_1 - c}{\sigma_1}), \\ \text{TPF}(c) &= 1 - \Phi(\frac{c - \mu_2}{\sigma_2}) = \Phi(\frac{\mu_2 - c}{\sigma_2}), \qquad (2) \end{aligned}$$

where Φ is the standard normal cumulative distribution function. Therefore,

$$\Phi^{-1}(\text{TPF}) = \frac{|\mu_2 - \mu_1|}{\sigma_2} + \frac{\sigma_1}{\sigma_2}\Phi^{-1}(\text{FPF}). \qquad (3)$$

To apply this method to evaluate the classification performance of CAD algorithms or an imaging modality. We first estimate the parameters: means and variances from the sample means and variances:

$$\hat{\mu}_j = \frac{1}{N_j}\sum_{i=1}^{N_j} x_{ij}, \hat{\sigma}_j^2 = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ij} - \hat{\mu}_j)^2, \qquad (4)$$

where j is either 1 or 2, representing object and background, N_j the number of pixels in the jth class, and x_{ij} the ith pixel belongs to class j of decision variable x. Then, we use equation 3 to plot smooth ROC curves.

3. THE PROPOSED NEW ROC CURVE BASED ON CONDITIONAL PROBABILISTIC MODEL

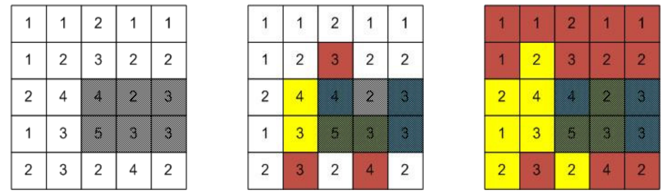
In the conventional ROC analysis, the intensity value of each pixel x_i is considered to be independent with those of other pixels in the image. However, there is a considerable correlation between x_i and its neighbors in real world applications such as medical imaging. In this paper, we propose to develop a new ROC method by defining the operating points as conditional FPF and TPF. The conditional FPF and TPF are the FPF and TPF given that the neighboring pixels belong to the object or background, as TPF|N, FPF|N, TPF|P and FPF|P. Here, FPF|N is defined by the fraction of actually background pixels incorrectly classified as "object" given that the neighboring pixels are classified as background, TPF|N is defined by the fraction of actually object pixels correctly classified as "object" given that the neighboring pixels are classified as background. Similarly, FPF|P is defined by the fraction of actually background pixels incorrectly classified as "object" given that the neighboring pixels are classified as object, TPF|P is defined by the fraction of actually object pixels correctly classified as "object" given that the neighboring pixels are classified as object. In this way, the new ROC analysis is comprised of two curves, one is given the neighboring pixel classified as object (FPF|P versus TPF|P) and the other is given the neighboring pixel classified as background (FPF|N versus TPF|N). These conditional dependencies allow us to create ROC curves that incorporate the correlation between pixels.

A simple example shown in Fig.2 explains how to calculate the corresponding FPF|P, TPF|P, FPF|N and TPF|N with varying thresholds. For classical ROC analysis, we sort the intensity values of the example image (Fig.2(a)) as $\{1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,4,4,4,5\}$. By varying the threshold from 1 to 5, we get the FPFs and TPFs as $\{1,\frac{13}{19},\frac{10}{19},\frac{8}{19},0,0\}$,$\{1,1,\frac{5}{6},\frac{3}{6},\frac{1}{6},0\}$. To calculate FPF|P, TPF|P, FPF|N and TPF|N, we simply select the intensity value to the right pixel for each as the neighboring information, and pad the boundary by 0. Fig.2 demonstrates how the proposed method works. In Fig.2(b), we can see that when the threshold is 3, the pixels whose intensity values no less than 3 are classified as object. Red pixels are FP|N, because those pixels are incorrectly classified as positive given the right neighbor whose intensity value less than 3 is classified as negative. Yellow pixels are FP|P, because they are incorrectly classified as positive given the right neighbor is classified as positive. Blue pixels are TP|P, and they are correctly classified as positive given the right neighbor is also classified as positive. Green pixels are correctly classified as positive given the right neighbor is classified as negative, so they are TP|N. Similarly in Fig.2(c), red pixels are the negatives given the right neighbor is negative (N|N), the yellow pixels are the negatives given the right neighbor is positive (N|P), the blue pixels are the positives given the right neighbor is negative (P|N), and the green pixels are the positives given the right neighbor is positive (P|P). Therefore, for the example image, we can obtain the FPF|N and TPF|N as $\{1,\frac{3}{6},\frac{3}{12},\frac{1}{15},\frac{1}{18},0\}$,$\{1,1,1,\frac{2}{6},\frac{1}{6},0\}$ and the FPF|P and TPF|P as $\{1,\frac{10}{13},\frac{2}{7},\frac{1}{4},0,0\}$,$\{1,1,\frac{2}{3},0,0,0\}$. The new operating points are expressed as follows:

$$
\begin{aligned}
\text{FPF|N} &= \frac{\text{FP|N}}{\text{N|N}}, \qquad \text{TPF|N} = \frac{\text{TP|N}}{\text{P|N}} \\
\text{FPF|P} &= \frac{\text{FP|P}}{\text{N|P}}, \qquad \text{TPF|P} = \frac{\text{TP|P}}{\text{P|P}}.
\end{aligned} \tag{5}
$$

To fit the proposed conditional ROC curves, we assume the decision vector $y_i = [x_i \;\; n_i]$ belongs to a multivariate distribution. Then, we define the conditional TPF and FPF corresponding to varying threshold c as follows:

$$
\begin{aligned}
\text{FPF|N}(c) &= \text{Prob}(x > c | n < c, x \in S_1, n \in S_1) \\
&= \frac{\text{Prob}(x > c, n < c | x \in S_1, n \in S_1)}{\text{Prob}(n < c | n \in S_1)} \\
&= \frac{\int_c^{\infty} \int_{-\infty}^{c} p(x, n | x \in S_1, n \in S_1) dndx}{\int_{-\infty}^{c} p(n | n \in S_1) dn}, \\
\text{TPF|N}(c) &= \text{Prob}(x > c | n < c, x \in S_2, n \in S_2) \\
&= \frac{\text{Prob}(x > c, n < c | x \in S_2, n \in S_2)}{\text{Prob}(n < c | n \in S_2)} \\
&= \frac{\int_c^{\infty} \int_{-\infty}^{c} p(x, n | x \in S_2, n \in S_2) dndx}{\int_{-\infty}^{c} p(n | n \in S_2) dn},
\end{aligned} \tag{6}
$$



(a) (b) (c)

Fig. 2. Illustration of calculating an operating point with threshold = 3. Part (a) is an example image. Part (b), the red pixels are FP|N, the yellow pixels are FP|P, the blue pixels are TP|N, and the green pixels are TP|P. Part (c), the red pixels are N|N, the yellow pixels are N|P, the blue pixels are P|N, and the green pixels are P|P.

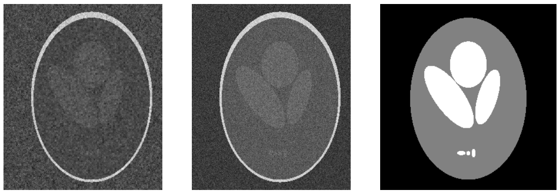
$$
\begin{aligned}
\text{FPF|P}(c) &= \text{Prob}(x > c | n > c, x \in S_1, n \in S_1) \\
&= \frac{\text{Prob}(x > c, n > c | x \in S_1, n \in S_1)}{\text{Prob}(n > c | n \in S_1)} \\
&= \frac{\int_c^{\infty} \int_c^{\infty} p(x, n | x \in S_1, n \in S_1) dndx}{\int_c^{\infty} p(n | n \in S_1) dn}, \\
\text{TPF|P}(c) &= \text{Prob}(x > c | n > c, x \in S_2, n \in S_2) \\
&= \frac{\text{Prob}(x > c, n > c | x \in S_2, n \in S_2)}{\text{Prob}(n > c | n \in S_2)} \\
&= \frac{\int_c^{\infty} \int_c^{\infty} p(x, n | x \in S_2, n \in S_2) dndx}{\int_c^{\infty} p(n | n \in S_2) dn}.
\end{aligned} \tag{7}
$$

By varying the threshold c, the proposed ROC curves are fitted to the conditional probabilities obtained by Eq. 7.

4. EXPERIMENTAL RESULTS

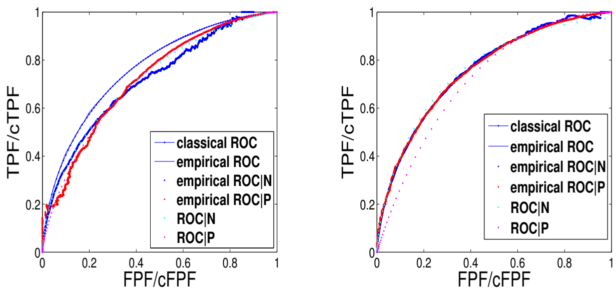
To demonstrate the usefulness of the proposed ROC curves, we apply the conventional and the proposed new ROC method to two computer simulated images. These images might represent both an imaging modality or a resulting image of a classification algorithm that fused images from multiple modalities. The simulated images consist of two regions: a background tissue and simulated organs or region of interests. The size of the images is 300×300. The intensities of the background and the object belong to Gaussian distributions. For all the experiments, we randomly select one of the 4-neighborhood pixel as the neighbor for each pixel. Fig.3 and 4 provide the example simulated images, classical ROC curves, and the proposed ROC curves.

In our experiments, we call the original ROC curve without modeling as empirical ROC curve, and the classical ROC curve is the binormal model to fit the empirical ROC curve by Eq.3. For the two simulated images, the means and variances of the background and object are the same. However,



(a) (b) (c)

Fig. 3. Simulated images and ground truth.



(a) (b)

Fig. 4. ROC curves with different models. Part (a) provides the ROC curves of image shown in Fig.3(a), and part (b) the ROC curves of image shown in Fig.3(b).

for the image on the left Fig. 3(a), the neighboring pixels are correlated whereas for the image on the right Fig. 3(b), the neighboring pixels are independent. There are three points we could observe from Fig.3 and 4:

(1) By visually evaluating the images, we can see that the image on the right outperforms the image on the left in terms of the classification of the background and object.

(2) The empirical ROC curves are very close and the classical ROC curves are identical because the means and variances of the two classes are the same in these two images. This observation is not consistent with the visual evaluation. **The distinguishability of the objects in the images are in fact different whereas classical ROC curves are the same for the two images.**

(3) **By comparing the conditional ROC curves, we can see that the new ROC curves for the image on the left is inferior to that of the image on the right.** This observation is consistent with a visual evaluation.

As a result, the proposed ROC analysis is able to quantify the classification power of an image much more accurately than classical ROC methods.

5. CONCLUSIONS AND DISCUSSION

ROC analysis plays a crucial role in performance evaluation of medical images and CAD systems. However, in the available ROC methods used to evaluate classification power, the pixels within the image are considered to be independent. The correlation between neighboring pixels are ignored. In this paper, we propose to consider the correlation between neighboring pixels and develop ROC methods based on conditional probabilistic model. We assume the intensity values of each pixel and that of its neighbors as a decision vector instead of assuming the intensity values of each pixel as a decision variable in the classical ROC analysis. With conditional probabilistic modeling, the ROC curves demonstrate the relationship between true positive fraction and false positive fraction given that the neighboring pixels are positive or negative. We have found that the new ROC curves measure the localization accuracy of a classification algorithm or an imaging modality more accurately than traditional ROC methods, since it takes into account the correlation between neighboring pixels.

6. REFERENCES

[1] A. Wald, "Statistical decision functions," Wiley, New York, 1950.

[2] J. P. Egan, "Signal detection theory and ROC analysis," Academic, New York, 1975.

[3] C. E. Metz, B. A. Herman, and J. Shen "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously distributed data," *Stat. in Medicine*, 17:1033-1053, 1998.

[4] R. G. Swensson, "Unified measurement of observer performance in detecting and localizing target objects on images," *Med Phys*, 23:1709-1725, 1996.

[5] D. P. Chakraborty "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Med Phys*, 16:561-568, 1989.

[6] L. M. Popescu "Model for the detection of signals in images with multiple suspicious locations," *Med Phys*, 12:5565-5574, 2008.

[7] L. Wei, Y. Yang and etc. "Relevance vector machine for automatic detection of clustered microcalcifications," *IEEE trans. on Med. Imaging*, 24(10):1278-1285, 2005.

[8] G. Langs, P. Peloschek, H. Bischof, F. Kainberger "Automatic quantification of joint space narrowing and erosion in rheumatoid arthritis", *IEEE trans. on Med. Imaging*, (28)1:151-164, 2009.