THE INFO-ROC TECHNIQUE: A METHOD FOR COMPARING AND OPTIMIZING INSPECTION SYSTEMS

Eugene Somoza[1,2], Douglas Mossman[3], and Larry McFeeters[2]

[1]Psychiatry Service, V.A. Medical Center
3200 Vine St. Cincinnati, OH 45220

[2]Division of Neuroscience, Dept. of Psychiatry, Univ. of Cincinnati College of Medicine, 234 Bethesda Ave. Cincinnati, OH 45267-0559

[3]Department of Psychiatry, University of Cincinnati College of Medicine, 234 Bethesda Ave. Cincinnati, OH 45267-0559

INTRODUCTION

Inspection systems are used in many areas of manufacturing to detect defects in machinery or components which might result in malfunction. Such defects often can be detected definitively by methods that establish "the truth" about whether defects are present, but these definitive methods suffer their own drawbacks: they generally are more complicated, more expensive, more time consuming and, most important of all, more destructive to the manufacturing system or components being evaluated.

In developing a non-destructive inspection system, it is necessary to assess and compare its accuracy in measuring defects with other detection systems and with the definitive methods alluded to above. In this paper we show how receiver operating characteristic (ROC) analysis describes the mathematical relationship between an inspection system and the relevant "gold standards". We also show how ROC analytic indices and information theory can be used to operationalize the inspection system for maximum efficacy; further, we show how information theory can be used to compare various inspection systems. We refer to this combination of ROC analysis and information theory as "the INFO-ROC Technique." For demonstration purposes, we here apply the technique to a specific inspection system, in which fluorescent penetrant dye is used to detect cracks in aircraft engine turbine blades.

CALIBRATING THE INSPECTION SYSTEM

The first step in the development of an inspection system is to assess it against the truth as defined by some gold

Review of Progress in Quantitative Nondestructive Evaluation, Vol. 9
Edited by D.O. Thompson and D.E. Chimenti
Plenum Press, New York, 1990

standard. In the present application, we began with data on 200 turbine blades known to have cracks and 655 known to be free of cracks. In this case, the true status of the blades was readily determined, in that the cracked blades had been damaged especially for the purpose of calibrating the inspection system; their status, and that of the uncracked blades, was confirmed by visual inspection. It was assumed that these calibration blades were similar to those that eventually would be tested using the inspection system.

The calibration blades were then evaluated using the inspection system being developed. Although we will not describe the process in detail, the blades are prepared by washing, applying fluorescent dye and developer, and washing again. Finally, the blades are stimulated with electromagnetic radiation of a certain frequency and the emitted fluorescent radiation is collected. The detection of cracks in the blades depends on the number of pixels of emitted radiation obtained. We will refer to this variable (i.e., number of pixels) as the inspection system variable (ISV). We note at this point, that in general, all systems involve the measurement of one or more ISVs for the parts being tested. In this paper we have limited ourselves to systems that use only one ISV.

Once the calibration parts have been evaluated with the inspection system, the ISV values may be plotted on a scatter graph such as that shown on the left side of Figure 1. The scatter graph shows the distribution of the ISV values obtained for the blades with no defects (open circles) and for the blades with defects (closed circles). The ordinate on this graph is the number of pixels. Note that the graphs of Figure 1 were made purely for illustration purposes, and do not utilize the actual experimental data that is used in subsequent figures.

From Figure 1, one can see that uncracked blades tend to yield fewer pixels than do cracked blades, but there is a

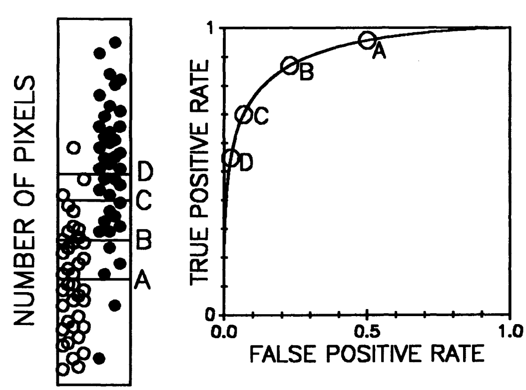

Fig. 1. Relationship between a scatter graph and the corresponding ROC curve for an inspection system.

large degree of overlap between the two sets of ISV values. The existence of this overlap leads to one of the fundamental problems in the development of an inspection system: How should one choose an appropriate cut-off, or (to put it in a different way) how to choose an operating point of the inspection system?

CONSTRUCTION OF A ROC CURVE

The problem of choosing an operating point for an inspection system is addressed by describing graphically or mathematically the trade-offs involved in choosing one cut-off over another. In the scatter graph of Figure 1, four possible cut-offs, labeled A through D, are shown. Note that when a certain cut-off is chosen, the two sets of blades (cracked and uncracked) are divided into four groups: the cracked blades lying above the cut-off are true positives; the uncracked blades lying above the cut-off are false positives; the cracked blades below the cut-off are false negatives; the uncracked blades below the cut-off are true negatives. From these four quantities one can define four terms that characterize the inspection system at a given cut-off. The true positive rate (TP, also known as the sensitivity) equals the number of true positives divided by the total number of defective blades; the false positive rate (FP, or "false alarm" rate) equals the number of false positives divided by the total number of non-defective blades. Two other quantities, the true negative rate (TN, or specificity) and the false negative rate (FN), can be similarly defined; also, the false negative rate equals 1-TP, and the true negative rate equals 1-FP.

An alternate way of defining TP and FP is the following. The TP can be defined as the conditional probability that, if the part is defective, then the inspection system will identify it as defective (i.e., $P(I^+|D^+)$). The FP can be defined as the conditional probability that, if a blade is non defective, the inspection system will identify it as defective (i.e., $P(I^+|D^-)$.

We wish to emphasize that TP, FP, TN and FN all depend on the choice of cut-off. If we look at the scatter diagram in Figure 1, we see that as the cut-off increases (i.e., from A to D), TP and FP both decrease whereas the TN and FN increase. The relationship between choice of cut-off and the values of TP and FP can be described graphically as shown on the right side of Figure 1. Here, for each of the four cut-offs, we have calculated the corresponding TP and FP and have plotted them in the figure. Many other points could be placed on this figure by choosing alternative cut-offs. The locus of all TP and FP pairs when plotted on a graph such as that of Figure 1, describes the "receiver operating characteristic" (ROC) of the inspection system. A ROC curve describes the relationship between true positive rate and false positive rate for <u>all</u> possible choices of cut-off in an inspection system [1].

The ROC curve shown in Figure 2 was obtained from actual fluorescent penetrant data from 855 turbine blades, 200 of which were cracked. The nine points shown on the ROC curve represent nine values of cut-off used to construct the curve. The smooth curve joining the curve has been generated by the method to be described below. In examining ROC curves such as Figure 2 one quickly learns that some correspond to more discriminating inspection systems than others. In general, the closer the ROC curve approaches the upper left hand corner of the unit square, the better the inspection system; important exceptions to this will be described below. From a quantitative point of view, the closer that the area under a ROC curve approaches 1.0, the better the corresponding inspection system is in discriminating between parts that are defective from parts that are not [2]. The area under the curve in Figure 2 is 0.958.

Up to now we have not made any <u>a priori</u> assumptions about the distributions of values of the ISV. For many diverse systems these distributions are normal (Gaussian) or can be made to approximate normal distributions by an appropriate monotonic transformation of the original data [3]. One can check the conformity of actual data with this "binormal assumption" using a chi-square test; we found that the data on turbine blades was consistent with binormality. That is, the transformed values of the ISV for the cracked and uncracked blades formed overlapping, normal distributions. These distributions may be described in terms of the difference between their means (Δm, in units of the standard deviation of the uncracked blades' ISV) and the ratio s of the standard deviations (s.d. uncracked/s.d. cracked) [4].

For data that conforms to the binormal assumption, a plot of (FP, TP) pairs on binormal coordinates (a plot of the normal deviates [Z_{FP}, Z_{TP}] on linear coordinates) will approximate a straight line, with slope = s and x-intercept = $-\Delta m$. The appropriate "best fit" technique uses maximum likelihood estimation; the available computer software calculates estimates of slope and intercept plus their

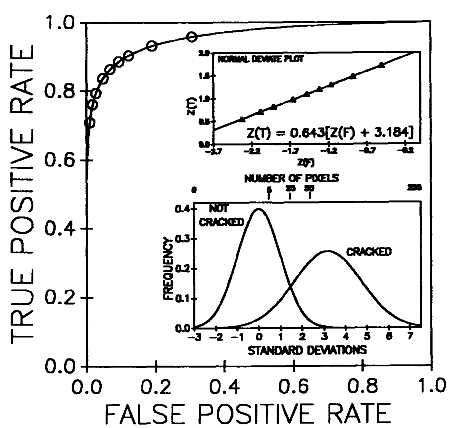

Fig. 2. ROC curve, normal deviate plot and binormal distribution for a fluorescent penetrant dye inspection system.

variances [5]. When this was done with the fluorescent penetrant data, the curve shown in the top insert of Figure 2 was obtained. The results for Δm and s are 3.184 and 0.643, respectively. With these two indices we used the linear equation shown on the top insert in Figure 2, to generate normal deviates of TP. These (Z_{FP}, Z_{TP}) pairs were then converted to (FP, TP) values by using algorithms for the normal distribution, allowing the smooth, continuous ROCcurve to be generated.

We see from the above that <u>any inspection system can be described</u>, <u>for all possible values of cut-off</u>, <u>by its ROC curve</u>; furthermore, if the raw or transformed data are normally distributed, then the ROC curve itself is <u>completely</u> described by the two indices Δm and s. These facts will be used below in developing techniques for operationalizing inspection systems in such a way to guarantee maximum information yield.

THE ROLE OF BAYES' THEOREM IN EVALUATING INSPECTION SYSTEMS

To this point, our discussion has focused on the description of intrinsic properties of an inspection system. Although a ROC curve describes the properties of an inspection system throughout its full range of possible cut-offs, it cannot, by itself, be used to predict how the inspection system will perform when used in the field. The reason for this is that the performance of an inspection system depends not only on the intrinsic properties of the inspection system itself but also on the particular circumstances in which it is used. A formal statement of this concept is known as Bayes' theorem; the "particular circumstances" we refer to are the prior likelihood for defects, or the prevalence [6].

One way of arriving at a qualitative understanding of Bayes' theorem is to consider the significance of the definition of TP and FP. TP is the conditional probability that, if a turbine blade actually is defective, it will be detected as defective by the inspection system. Though this is crucial information in the process of assessing and calibrating the inspection system, it is useless when the system is actually being used. At that point, one needs the inverse probability, i.e. the conditional probability that, if the inspection system determines that a turbine blade is defective, the blade indeed is defective. The relationship between the latter probability $P(D^+|I^+)$ and the former probability $P(I^+|D^+)$ involves not only the intrinsic properties of the inspection system but also the prevalence of defects in the blades being tested [6]. Blades made from a certain type of material or manufactured in a certain era may be more predisposed to having cracks than blades made from a different material or at a different time. The actual performance of an inspection system at a specific cut-off will vary depending on the prevalence of cracks in the class of blades being tested.

INFORMATION THEORY

ROC analysis by itself is not very helpful in determining which cut-off should be chosen for a given inspection system to be used in a given situation. It is also not very helpful in determining which of several inspection systems is best when used to test turbine blades with a known prior probability (prevalence) of defects. One way of addressing both of these problems is by means of information theory. Whenever an inspection system is used to evaluate an item, there is a gain in the amount of information known about that item. Using Shannon's definition of information, one can calculate the amount of information gained when an inspection system is used to test an item for defects if the prevalence of defective items is known. This information gain is a function of three variables, TP, FP and prevalence (P). The application of information to detection systems was first done by Metz in 1973 [7], who showed that, for every point on a ROC curve of a given detection system, one can calculate the information gained for a given value of prevalence.

When we graph information gained vs. the cut-off at a certain prevalence, we find that the curve always has a maximum. This means that there is a certain cut-off at which the information gained as a result of using the inspection system is a maximized. If there are no other <u>a priori</u> criteria for choosing a cut-off, this intrinsic criterion of maximizing information gained is a quite reasonable one to use for selecting a system's operating point.

Using the data on cracked and uncracked turbine blades discussed above, we calculated information as a function of cut-off for fifty values of prevalence between zero and one, and then found the information maximum from each such curve [8]. These values of maximum information were plotted as a function of prevalence in the main portion of Figure 3. The figure depicts the maximum information attainable for this inspection system, in bits, for each value of prevalence. For each value of prevalence this maximum information was obtained at a different cut-off.

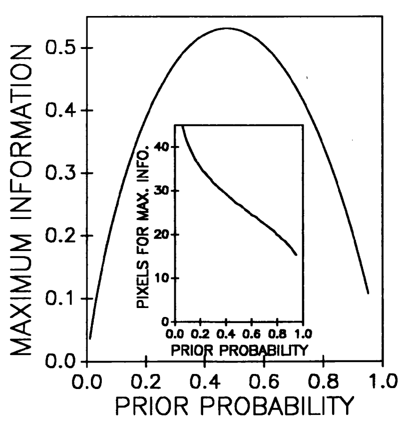

Fig. 3. **Maximum information vs. prevalence (MIP) and cut-off for maximum information vs. prevalence (CMI) curves.**

The maximum information vs. prevalence (MIP) curve depicts the effectiveness of the inspection system as a function of the prior probability of cracks in the turbine blades being tested. The graph's horizontal and vertical axes do not involve units that are related to the specific nature of the inspection system, which means that many different inspection systems' MIP curves can be compared with each other even if they arise from totally different technologies (e.g., the eddy current method vs. the fluorescent penetrant dye method for detecting cracks). Inspection systems whose MIP curves indicate a higher information yield would be superior to those that yield lower values of information. It is possible for some inspection systems to do better than others over a certain range of prevalence, but worse over a different prevalence range [9].

One can calculate and graph cut-offs for maximum information in a manner similar to that used to generate MIP curves. At any prevalence, one can determine which cut-off yields maximum information; one can also plot a set of these cut-offs for a range of prevalence values. This was done to create the insert to Figure 3 [8]. For a given prevalence, this curve (the cut-off for maximum information, or "CMI" curve) can be used to determine the optimum operating point of the inspection system.

DISCUSSION

The INFO-ROC technique can be used to carry out three important analytical processes with regard to inspection systems: 1) to describe them mathematically for all possible values of cut-off; 2) to find the optimal cut-off at a given prior probability; and 3) to compare various inspection systems directly with each other as a function of prior probability.

This method takes into consideration the important concept that if the cut-off is varied in such a way that the false positives decrease, it will also cause the false negatives to increase. This trade-off between the false positive rate and the false negative rate is not taken into consideration in the other method currently being used to evaluate inspection systems, viz., the "probability of detection" method. This method ignores the existence of false negatives.

The INFO-ROC Technique also utilizes the Bayesian concept that the performance of any inspection system depends not only on the intrinsic properties of that system, but also on the prior probability of faults in the material being inspected. Thus, it is not possible to choose a cut-off for a given inspection system and expect it to be the best cut-off possible for any set of material to be tested. The cut-off must be individualized to reflect the prior probability of defects in the items to be evaluated.

In this paper we have assumed that maximizing the information gained is a desirable criterion and can be used to operationalize the inspection system. This is particularly useful in the process of developing inspection systems. For example, in the fluorescent penetrant dye system we discussed, one can test different dyes or different developers or different electromagnetic frequencies and see how these variables affect the MIP curves when the system is tested against standard turbine blades. Clearly, in such a situation, one would choose that version of the inspection system that gives the highest MIP curves. However, there are other situations where information alone is not the only criterion that should be used. There are often questions of costs, benefits, and risks which may also be taken into consideration in the choice of the cut-off. We have combined cost-benefit analysis with ROC analysis for a different detection system; this study will be published elsewhere [10].

Finally we would like to address the important issue of the distribution of values of the ISV. As mentioned above, there are many circumstances in which either the ISV values themselves or a transformation of them closely approximates the normal distribution (see bottom insert of Figure 2), and this can be evaluated using appropriate goodness-of-fit tests [11]. There may be some situations, however, where the binormal assumption is invalid. In such cases, one must develop a different relationship between the two variables of the ROC curve (TP and FP) in order to carry out the subsequent steps of calculating information gain from the inspection system. All the steps described for the INFO-ROC Technique can then be used after the functional relationship between FP and TP is known.

REFERENCES

1. J.A. Swets, Science, 240:1285-1293 (1986).
2. J.A. Hanely and B.J. McNeil, Radiology, 143:29-36 (1982).
3. J.A. Swets, Psychological Bulletin, 99:181-198 (1986).
4. C.E. Metz, Seminars in Nuclear Medicine, 8:283-298 (1978).
5. C.E. Metz, H.B. Kronman, P.L. Wang, J.H. Shen, ROCFIT: A modified Maximum likelihood algorithm for estimating a binormal ROC curve Confidence rating data. (University of Chicago, Chicago, 1978).
6. B.J.McNeil, E. Keeler, S.J. Adelstein, New England Journal of Medicine, 293:211-215 (1975).
7. C.E. Metz, D.J. Goodengough, K. Rossmann, Radiology, 109:297-304 (1973).
8. E. Somoza, L. Soutullo-Esperon, D. Mossman, International Journal of Biomedical Computing. (In press).
9. E. Somoza and D. Mossman, Biological Psychiatry, (In press).
10. E. Somoza and D. Mossman, In preparation.
11. C.E. Metz, Investigative Radiology, (In press).