Quantitative assessment of radiographic control informativeness using ROC analysis

A S Sorokin¹, D I Galkin², E A Ivanayskiy³ and A E Shubochkin²

¹ Plekhanov Russian University of Economics, Moscow, Russia,
² CJSC RII MSIA "Spektrum", Moscow, Russia
³ Polzunov Altai State Technical University, Barnaul, Russia

Abstract In industrial radiography the issue of replacment film detectors by digital ones becomes increasingly relevant. Also remains unresolved the question of an objective comparison of various radiographic testing (RT). This article presents the results of research, aimed at obtaining a quantitative assessment of RT technology informativeness. In order to construct ROC curves, characterizing the quality of binary classification, using a specific RT technology, there was designed and manufactured a sample test, containing a simulation of the most difficult to detect discontinuities. Such test sample was made using additive technology. Developed manufacturing technology made it possible to ensure sufficient accuracy of linear dimensions, taking into account the product's design features. After exposure of test object to various detectors, there was a decoding procedure, conducted by experts, whose task was to establish the presence / absence of a defect in analyzed element of the image. Applying such technique, used in results analysis of deciphering the images of sample test, made it possible to minimize the influence of human factor and obtain ROC curves, reflecting the capabilities of RT technology only. The subsequent determination of ROC curves parameters allows to conduct a comparative analysis of informativeness of the technologies under consideration.

1. Introduction

Today, the issue of replacing the film technology of radiographic control (RC) with digital is becoming increasingly important [1]. Thanks to the latest CCD and CMOS matrices, more and more digital detectors appear on the nondestructive testing (NDT) market [2]. Normative documents regulating the requirements to the technology of radiographic control with the use of phosphoric plates (computer radiography) and matrix detectors (digital radiography) are introduced. However, the peculiarities of optical image formation when using digital systems (MTF modulation transfer function) lead to gradation image alignment, which can result in the "disappearance" of some images after their software processing and presenting on the output device and the loss of useful information in comparison with film radiography.

It is obvious that the replacement of the film image with a digital one is possible in case of achieved detectability (informativeness of the image) corresponding to the radiographic image is achieved. In practice, informativeness means the amount quantity of useful information that can be extracted during the decoding process.
The aim of the present work is to develop the methodology to determine the quantitative indicator of informativeness for the subsequent comparison of RC technologies and the possibility of their application for the RC of a specific object.

The task of the RC specialist is to classify the image into two possible categories: defect or background. ROC-analysis (Receiver Operating Characteristic) is used to assess the quality of such a binary classification problem. The term ROC was first introduced during World War II, after the defeat of the US Navy at Pearl Harbor in 1941, when the problem of improvement for recognition of enemy aircraft by radar signal was realized. Today, the ROC-analysis method is widely used in credit scoring, medical diagnostics, customer loyalty prediction, acceptance quality control and other areas [3]. There are some attempts to use ROC-analysis in nondestructive testing [4, 5].

Let's consider the problem of classification from the point of view of RC. Figure 1 shows the probability densities of normal distribution of optical density of the image corresponding to two categories – defect and background. The applied RC technology will influence on the mutual position and form of the distributions of both categories. The intersection of these graphs leads to uncertainty for decision-making. The following types of solutions may appear as a result of classification:

- **True Positive (TP)** – the defect is recognized as a defect;
- **True Negative (TN)** – background is recognized as background;
- **False Positive (FP)** – type I error, the background is recognized as a defect or "rejection»;
- **False Negative (FN)** – type II error, defect is recognized as background or "defect".

![Figure 1. Explanation to the construction of ROC-curve.](image)

The probability of making decisions of each type is determined by the cut-off threshold position, fixing the boundary value of the optical density, the excess of which means for the classifier that the object belongs to the "defect" category. In the opposite side - the "background" category. Cut-off threshold characterizes the individual characteristics of the classifier (RC specialist). As shown in Figure 1, if the threshold is at point D, then the proportion of false negative results will be higher; at point C, on the contrary, the proportion of false positive cases will be higher. And at point E, there will be a balance between type I and type II errors.

The basic concepts in the quality analysis of binary classifiers are:

- **sensitivity** – the proportion of true positive decisions;
- **specificity** – the proportion of true negative decisions;
- **accuracy** – the total proportion of correct decisions;
- **the area under the curve (AUC)** – the area under the ROC-curve, one can use an alternative indicator of Gini index, which is $2\text{AUC} – 1$;
- **cut-off point** – value for decision making, for example, the point where the highest sensitivity and specificity value is reached;
- **Yoden index** – includes sensitivity and specificity at the same time: the sum of sensitivity and specificity minus 1, it reflects the forecast balance.

ROC-analysis is a graphical representation of the proportion of true positive and false positive results from the selected cut-off point. If the ROC-curve is a straight line from the origin point to the upper right corner, it corresponds to the absence of reliable data on the presence of the defect (the
probability densities of the defect and the background coincide). The points on this line indicate that the test gives the same number of truly positive and false positive results, and therefore does not distinguish between areas with or without defect. The higher the ROC-curve to the upper left corner of the coordinates and, accordingly, the larger area under the ROC curve (AUC), the higher the information content of the diagnosed RC technology. And the optimal clipping point on the ROC-curve will be as close as possible to the upper left corner of the coordinate system.

ROC-analysis method can be used in RC to solve the following problems:

- assessment of the classifier quality: value of technology informativeness, the objective of qualification assessment;
- determination of the optimal cut-off point: reasonable selection of rejection rates for a particular control technology, application in the creation of machine vision algorithms;
- risk assessment of incorrect classification during application of specific control technology;
- classifiers comparison: comparative analysis of the capabilities of different RC technologies.

2. Planning and conducting of the experiment

2.1. Experiment design and sample manufacturing
To assess the RC method informativeness for constructing ROC-curves the test object was made (Figure 2) – steel plate 100×100 mm thickness 10 mm, divided into 100 identical square slots.

Figure 2. Test sample photo.

One separate square, could contain not more than one defect simulator – a short rectangular groove with an opening width of 0.1 mm; the grooves depth is ranged from 0.1 to 0.5 mm (sampling step is 0.1 mm). According to GOST 7512-82, during RC, it is necessary to ensure the defects detection with this minimum opening width, i.e. during the experiment the most difficult case for detection defects was simulated. The number of slots, their coordinates and depth were determined randomly.

The test sample was manufactured using SLM printing technology (laser melting) on 3D printer model SLM 280HL using YLR-fibre-laser, 400 watts power. Surfacing was carried out by 0.03 mm thickness layers. Printing was maid in nitrogen atmosphere using steel spherical powder 316L with particle size of 0.000045…0.00015 mm. Modeling and designing of test sample was performed using SLM Build Processor, Magics RP, Magics SG+ softwares.

2.2 Data collection for information content evaluation
The statistical data collection to assess the achieved informativeness in the application of particular RC technology was carried out by the expert survey. Picture-image of the test sample (Figure 3) was presented to assess the discontinuity presence to RC experts. Their task was to match each of these 100 elements of the image with one of the categories, given in the Table 1.
Figure 3. The image of the test sample obtained using the digital detector "PRODIS mark 1215SS" (scintillator CsI, pixel size 50 microns, resolution of frequency-contrast characteristic – 100 microns).

Table 1. Categories of expert solutions during evaluation of each sample slot.

| No. | Solution category                                                                 | The probability of the discontinuity projection presence |
|-----|----------------------------------------------------------------------------------|---------------------------------------------------------|
| 1   | The slot contains the discontinuity image                                         | 1                                                       |
| 2   | The slot most likely has the discontinuity image and I will make the record about it in the Protocol | 0.75                                                   |
| 3   | Additional research is required to make the decision                              | 0.5                                                    |
| 4   | In the slot most likely does not have discontinuity image, and I will not fix it in the Protocol | 0.25                                                   |
| 5   | In the slot does not have discontinuity image                                     | 0                                                      |

According to the decoding results one expert compiled the table containing the dependence of responses of the category values and the actual presence of discontinuity in the sample slot (Table. 2).

Table 2. Example of solution category values for a single expert.

| Actual state | Number of responses in each solution category | Total |
|--------------|----------------------------------------------|-------|
|              | 1          | 0.75  | 0.5 | 0.25 | 0   |       |
| Discontinuity| 21         | 9     | 0   | 4   | 14  | 48    |
| No discontinuity | 1         | 0     | 0   | 9   | 42  | 52    |

2.3. Method of particular ROC-curve constructing
Table 2 is the basis for ROC-curve calculating for the specific expert (particular ROC-curve). In order to construct curve points, it is necessary to calculate their coordinates. In this case, the ordinate axis corresponds to the probability of truly positive solutions, the abscissa axis - to the probability of false positive solutions.
To construct the first point the maximum confidence solutions (category with probability estimation 1) related to the total number of cells in this row are used. The original point of ROC-curve graph for ordinate axis - 21/48 = 0.44, for abscissa axis - 1/52 = 0.02.

The points coordinates for other decisions categories are determined in the same way. Each subsequent table cell is calculated as the sum of the current category and all previous ones. The value of the 2nd category (0.75) is calculated as the sum of the 1st and 2nd; the 3rd category is the sum of the 1st, 2nd and 3rd, etc. The result is a table of all points values for ROC-curve construction (Table 3 and Figure 4).

Table 3. Finding of the points coordinates for ROC-curve construction.

| Axis | Coordinate |
|------|------------|
| y    | 0 0.44 0.63 0.63 0.71 1.00 |
| x    | 0 0.02 0.02 0.02 0.19 1.00 |

Figure 4. ROC-curve constructed during the decoding of the test sample image by the specific expert.

To determine AUC as quantitative criterion for detecting discontinuities, it is necessary to sum up the areas on the separate segments \([x_i; x_{i+1}]\):

\[
S_i = \frac{(y_i+y_{i+1})(x_{i+1}-x_i)}{2}
\]

(1)

The area under the AUC varies from 0.5 (no informativeness) to 1.0 (maximum informativeness). The bigger the square value, the more informative the control technology during the technology assessment or the higher expert qualification in the qualification assessment. The square value 0.9...1.0 corresponds to the excellent classifier quality, 0.8...0.9 – very good, 0.7...0.8 – good, 0.6...0.7 – average, 0.5...0.6 – unsatisfactory. On Figure 5 the square value was 0.809.

The manual ROC-curve construction method is considered above. For automatic ROC-curves construction using primary data it is more efficient and expedient to use statistical packages of commercial and free applications (SPSS, Stata, R, etc.).
2.4. Estimation of the total ROC-curve

The ROC-curve constructed in this way reflects the opinion of one expert and cannot be objective. In order to make experiment statistically significant, it is necessary to minimize the influence of the human factor by determining sufficient number of independent experts.

To this purpose, 32 independent respondents decoded the image. To exclude psychophysical factors, respondents received the images rotated in different ways. For each expert it was recorded: age, presence/absence of qualifications, presence/absence of glasses. ROC-curves were constructed and square value of AUC were calculated for all 3200 observations and for each individual respondent. Figure 5 shows the ROC-curve for all respondents. All calculations were performed in IBM SPSS Statistics 24 (IBM). The value of AUC 0.742 (95% confidence interval from 0.725 to 0.760) was received.

![Figure 5. ROC-curve constructed according to 32 experts data.](image)

The AUC particular values for each experts were analyzed separately. Although the sample of only 32 respondents, the distribution of AUC value is very close to normal: average value is 0.74, standard deviation is 0.047, median is 0.748, minimum - 0.623, maximum - 0.824. The hypothesis of distribution normality tested by Shapiro–Wilk criterion is not rejected (p = 0.18). 95% confidence interval for the average AUC was from 0.727 to 0.761. These results practically coincided with the results of AUC construction for all observations.

The influence of psychophysical factors on the AUC value was also analyzed. Spearman's rank correlation coefficient between age and AUC is 0.09 and it is insignificant. The average AUC value scores in the group with and without qualification, glasses differ statistically slightly by t-criterion (p-values are 0.484 and 0.437, respectively). It shows that there were no influence of these factors on the decoding result.

2.5. Determining the minimum number of experts

On one hand, the decoding results obtained by individual expert will depend significantly on his qualifications. On other hand, the involvement of many experts to decoder the images of the test sample images is not always possible and requires significant time and finance. In this regard, the question arises: what is the minimum number of experts needed for test sample decoding consisting of 100 slots to ensure the results reliability, i.e. no influence of the human factor on the results. From mathematical point of view, the minimum number of experts should be determined.
According to the decoding results performed by 32 respondents, the AUC estimate distribution was close to normal, so the formula for determining the minimum number can be obtained from the formula for constructing the interval estimate for the average value:

\[
n = \frac{z^2 \sigma^2}{\Delta^2}
\]

where \( z \) is the value of the standard normal distribution law for given confidence probability \( \gamma \); \( \sigma^2 \) is the variance of the estimated index of experts; \( \Delta \) is the allowed limited absolute error in the evaluation of the average one. For this case, if we choose reliability \( \gamma = 0.90 \) (\( z = 1.64 \)), \( \Delta = 0.037 \) (this corresponds to the maximum error of 5% in the average AUC evaluation), the minimum number of respondents will be 5.

Thus, depending on the required reliability and accuracy as well as available resources, the formula allows to calculate the minimum number of experts for specific task for comparing the RC technologies, if the sample will have 100 slots for decoding. Thus, with reliability of 0.95 and limit error of 5%, 7 respondents will be enough. For 0.99 and 3%, the minimum number is 30 experts.

2.6. Example of method usage

As an example, let’s compare radiographic films Agfa D7 and Agfa D4 on the informative parameter. Characteristics of the films type, declared by the manufacturer, are shown in Table 4.

**Table 4.** Technical characteristics of Agfa D4 and D7 radiographic films for automatic chemical and photographic processing.

| Film type | Dose, Gr | Contrast ratio (gradient) | Granularity factor | Signal to noise ratio | Class according to ISO 11699-1 |
|-----------|-----------|---------------------------|--------------------|-----------------------|----------------------------------|
| Agfa D4   | 100       | 4.7                       | 9                  | 0.023                 | 232                              | 3                               |
| Agfa D7   | 320       | 4.5                       | 7.1                | 0.032                 | 142                              | 5                               |

\(^a\)\( D_0 \) – optical density of the veil.

More "beneficial" characteristics of Agfa D4 radiographic film in practical terms should have an impact on greater (compared to Agfa D7) detectability of discontinuities (due to larger gradient) with fewer rejections (due to higher signal/noise ratio). This dependence can be traced by analyzing ROC-curves constructed according to the authors’ method.

Since the objective of the study was to compare the film detectors, exposure modes shown in Table 5, were selected to ensure minimum geometric Unsharp, and the received images had the same optical density.

**Table 5.** Exposure modes and obtained image parameters.

| Film type | Type x/m | F, mm | U, kV | I, mA | \( t_{ckp}, \) c | f, mm | Snapshot options | Sensitivity K, mm | D |
|-----------|----------|-------|-------|-------|-----------------|-------|------------------|------------------|---|
| D4        | РАП1     | 2.0   | 180   | 5     | 260             | 800   | 0.1              | 2.4              |
| D7        | 220-5    |       | 180   | 3.5   | 150             | 800   | 0.1              | 2.4              |

\( x/m \) – x-ray machine; \( F \) – size of focal spot; \( U \) – voltage on x-ray tube; \( I \) – tube current; \( t_{ckp} \) – exposure time; \( f \) – distance "radiation source - object of control"; \( D \) – optical density of image.

The data obtained by images decoding of the test sample made by five experts was used to construct ROC-curves. Elements of the test sample with 0.5 mm grooves (100% detectability for these
RC technologies) and 0.1 mm – 100% pass were excluded from the samples. The number of observations to estimate the ROC-curve of D4 film was 328, and for D7 – 362.

Averaged over five measurements of ROC-curves corresponding to the considered RC technologies are shown in Figure 6. The area under the ROC-curve D4 was 0.77, and under D7 - 0.70. However, the area under the ROC-curve is not the only model quality indicator. The sensitivity and specificity analysis of the classifier model showed that the ROC-curve for D4 film is located closer to the upper left corner of the coordinates. At the nearest point to it for film D4 sensitivity is 65 %, and specificity – 88 % (Iodine index 0.52). And for D7 film, similar indicators will be 55 and 95%, respectively (Iodine index is 0.5). Thus, it can be concluded that the decision-making model based on D7 film is more balanced and gives sensitivity (the ability to see discontinuities) 10 percentage points higher.

![Figure 6. ROC-curves constructed for RC technologies using Agfa D7(a) and Agfa D4(b) radiographic film.](image)

3. Conclusion
The use of additive technologies aloud to produce the test sample with simulated defects, which can be used to assess the information content of radiographic control. The developed method of results analysis of the test sample decoding allows to obtain reliable results, substantially excluding the influence of human factor. The application of this method allows to obtain quantitative assessment of the information content of the considered radiographic control technology.

References
[1] Shinoda K 2012 Introduction of Digital Radiography Technique Jnl. of Japan Welding Society 81 (4) 244-250
[2] Shinoda K et al 2004 Application of Digital Radiography to Aerospace-Craft Key Engineering Materials 270-273 1361-5
[3] Georgiou G A 2006 Research Report 454 - Probability of Detection (PoD) curves Tech. rep. Jacobi Consulting Limited
[4] Liu C, Dobson J, Cawley P 2017 Efficient generation of receiver operating characteristics for the evaluation of damage detection in practical structural health monitoring applications Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 473 ISSN: 0080-4630
[5] Rajesh S N 2017 Performance evaluation of a magnetic field measurement NDE technique using a model assisted Probability of Detection framework NDT & E International 91 61-70