Report Number: WUCS-87-31

1987-11-01

An Evaluation of The Effectiveness of Adaptive Histogram Equalization for Contrast Enhancement

John B. Zimmerman, Stephen M. Pizer, Edward V. Staab, J. Randolph Perry, William McCartney, and Bradley C. Brenton

Adaptive Histogram Equalization (AHE), a method of contrast enhancement which is sensitive to local spatial information in an image, has been proposed as a solution to the problem of the inability of ordinary display devices to depict the full dynamic intensity range in some medical images. This method is automatic, reproducible, and simultaneously displays most of the information contained in the grey-scale contrast of the image. However, it has not been known whether the use of AHE causes the loss of diagnostic information relative to the commonly-used method intensity windowing. In the current work, AHE and intensity windowing are... **Read complete abstract on page 2.**

An Evaluation of The Effectiveness of Adaptive Histogram Equalization for Contrast Enhancement

John B. Zimmerman, Stephen M. Pizer, Edward V. Staab, J. Randolph Perry, William McCartney, and Bradley C. Brenton

Complete Abstract:

Adaptive Histogram Equalization (AHE), a method of contrast enhancement which is sensitive to local spatial information in an image, has been proposed as a solution to the problem of the inability of ordinary display devices to depict the full dynamic intensity range in some medical images. This method is automatic, reproducible, and simultaneously displays most of the information contained in the grey-scale contrast of the image. However, it has not been known whether the use of AHE causes the loss of diagnostic information relative to the commonly-used method intensity windowing. In the current work, AHE and intensity windowing are compared using psychophysical observer studies. In studies performed at North Carolina Memorial Hospital, experienced radiologists were shown clinical CT images of the chest. Into some of the images, appropriate artificial lesion were introduced; the physicians were then shown the images processed with both AHE and intensity windowing. They were asked to assess the probability that as given image contained the artificial lesion, and their accurate was measured. The results of these experiments shown that for this particular diagnostic task, there was no significant difference in the ability of the two methods to depict luminance contrast; thus, further evaluation of AHE using controlled clinical trials is indicated.

AN EVALUATION OF THE EFFECTIVENESS OF ADAPTIVE HISTOGRAM EQUALIZATION FOR CONTRAST ENHANCEMENT

John B. Zimmerman, Stephen M. Pizer,
Edward V. Staab, J. Randolph Perry,
William McCartney and Bradley C. Brenton

WUCS-87-31

November 1987

Department of Computer Science
Washington University
Campus Box 1045
One Brookings Drive
Saint Louis, MO 63130-4899

Submitted to IEEE Transactions on Medical Imaging

Abstract

Adaptive Histogram Equalization (AHE), a method of contrast enhancement which is sensitive to local spatial information in an image, has been proposed as a solution to the problem of the inability of ordinary display devices to depict the full dynamic intensity range in some medical images. This method is automatic, reproducible, and simultaneously displays most of the information contained in the grey-scale contrast of the image. However, it has not been known whether the use of AHE causes the loss of diagnostic information relative to the commonly-used method of intensity windowing. In the current work, AHE and intensity windowing are compared using psychophysical observer studies.

In studies performed at North Carolina Memorial Hospital, experienced radiologists were shown clinical CT images of the chest. Into some of the images, appropriate artificial lesions were introduced; the physicians were then shown the images processed with both AHE and intensity windowing. They were asked to assess the probability that a given image contained the artificial lesion, and their accuracy was measured. The results of these experiments showed that for this particular diagnostic task, there was no significant difference in the ability of the two methods to depict luminance contrast; thus, further evaluation of AHE using controlled clinical trials is indicated.

Index terms: AHE, psychophysics, contrast-enhancement, windowing

1 Introduction

Modern imaging devices such as computed tomography (CT) and magnetic resonance (MR) scanners provide images with a large range of contrast information. Unfortunately, the range of displayable intensities available on common display devices such as CRTs or film is insufficient to display all this information simultaneously. The usual solution to this problem is to apply a contrast enhancement mapping such as intensity windowing (or more precisely, global linear min-max windowing, to be defined hereafter) to the image to display the information of interest. This approach has the difficulty that it throws away much of the available information; with intensity windowing it is often necessary to apply several windows to the data in order to see the information available in different intensity ranges of the image. Further, intensity windowing requires manual intervention by the physician or technician to produce satisfactory results; often this manipulation must be repeated when the original film set is inadequate for complete interpretation.

An alternative method for contrast enhancement has been proposed by Pizer *et.al.* [1] which has the property that it is adaptive to *local* information in the image. This method, Adaptive Histogram Equalization (AHE), allows information in all intensity ranges of the image to be viewed simultaneously. It also has the advantages that it is automatic (*i.e.*, no manual intervention is required) and is thus reproducible from study to study. Preliminary evaluation of AHE has shown that it has promise for routine use in a clinical setting; however, before this can be realized, it must be demonstrated that AHE does not cause a loss of diagnostic information relative to intensity windowing in most circumstances.

As a first step in this demonstration, this paper describes psychophysical observer experiments which assess the relative effectiveness of AHE and global linear min-max windowing in allowing the detection of contrast while using grey-scale display for a particular diagnostic task.

2 Description of the Two Contrast Enhancement Modalities

In the current studies, Adaptive Histogram Equalization was compared with the most common clinical method of contrast enhancement, global linear min-max windowing. In this section these two contrast enhancement methods are briefly described.

Due to the coherence of objects within an image, the information about an object is generally contained in some subrange of the intensity values in the image. Global linear min-max windowing attempts to exploit this object coherence by enhancing the contrast within a subrange of the intensity values at the expense of the remaining intensity values. As shown in Figure 1, a subrange or window of the data is chosen and these intensity values are linearly remapped to occupy the entire range of the display device. A single such mapping is applied to the entire image. Intensity values lying above or below the window are mapped to the maximum and minimum possible intensity values respectively. The result is that the contrast of pixel values within the window is enhanced while information in other parts of the image is discarded. If other areas of the image are to be visualized, a new window must be chosen. Frequently, multiple windows are necessary to see all the information of interest in a clinical image. While standard windows can be selected to show particular areas of interest for a given anatomical region and imaging modality, manual intervention is usually necessary for the best results.

Adaptive Histogram Equalization attempts to overcome the limitations of global linear min-max windowing by providing most of the desired information in a single image which can be produced without manual intervention. In this technique, developed independently by Ketcham *et.al.*, Hummel, and Pizer *et.al.* [2-4], the contrast enhancement mapping applied to a particular pixel is a function of the intensity values immediately surrounding the pixel. The basic method is illustrated in Figure 2. For each pixel in the image, a region centered about the pixel, called its *contextual region* is assigned. The intensity values in that region are used to calculate a histogram equalization mapping which is then applied to the pixel in question. The result is an image in which the mapping applied to each pixel is different and is adaptive to the local distribution of pixel intensities rather than

the global information content of the image. In practice, this produces an image in which different objects whose intensity values lie in different subranges of the intensity values are simultaneously visible.

In the method just described, the mapping for each pixel is determined by calculating a histogram equalization based on the pixel's contextual region. This method works well, but is too slow for implementation on a general-purpose computer, requiring about two hours to process a 512×512 image on a VAX11/780 computer. An alternative algorithm approximates the mapping required for each pixel by choosing a small number of contextual regions within the image and calculating the mapping for a given pixel as a bilinear interpolation of the mappings derived from nearby contextual regions. This *interpolated* version of AHE gives results which are very close to the uninterpolated version. A full description of this method may be found in Pizer [4]. Interpolated AHE allows 512×512 medical images to be processed in a few tens of seconds on a general purpose minicomputer or a few seconds on many special-purpose image processing systems.

An example of interpolated AHE is shown in Figure 3. Here, an original chest CT image is shown along with two windowed versions, one intended to depict the lung area and the other the mediastinum, and the same image processed with AHE. It is clear that AHE allows the simultaneous visualization of the major vessels in the mediastinum and the pulmonary vessels.

3 Experimental Methodology

In order to evaluate the effectiveness of the two contrast enhancement methods described in the last section, formal observer studies were undertaken at North Carolina Memorial Hospital. The goal of these studies was to evaluate the methodologies using observer tasks which were as clinically realistic as possible while still allowing the results to be analysed by well-understood statistical methods. In this section, the experimental methodology which was used is presented.

3.1 Overview of the Experiment

A Receiver Operating Characteristic (ROC) rating experiment was performed to compare the ability of AHE and interactive global linear min-max windowing to depict luminance contrast. In this experiment, a set of test images was prepared from a number of normal CT images of the chest. In each normal image, four sites were chosen for the insertion of simulated lesions, two sites in the lungs and two in the mediastinum. For each field (lungs and mediastinum), three simulated lesions were prepared. The lesions varied in their linear size and grey-scale intensity. The linear sizes were chosen as appropriate for the given field; the intensities were chosen such that when a lesion was inserted into an image it was very subtle. From each normal image, twelve additional images were generated, each image having one of the lesions inserted at one of the selected sites. From each of these images, another image was prepared by processing the given image with AHE. The complete test set consisted of the normal images with no processing applied, the normal images with the lesions inserted and no processing applied, the normal images processed with AHE, and the normal images with the lesions inserted processed with AHE. In order that there be equal numbers of images with and without the artificial lesions, copies of the images without lesions inserted were included. The test set was then presented in random order to three trained radiologists.

For those images not processed with AHE, the observers were asked to perform interactive linear min-max windowing on the appropriate image field (lungs or mediastinum). After they had completed the windowing task, they were presented with a replica of the lesion placed to the side of the image and the exact location of the prospective lesion was indicated with a removable crosshairs. For those images processed with AHE, the observers were shown the lesion replica and the crosshairs immediately. In all cases, they were asked to rate their confidence that a lesion of the type depicted was present at the indicated site.

The task presented to the observers was thus one of detection of luminance contrast; the data constituted the results of a rating scale experiment from which a Receiver Operating Characteristic curve could be derived, giving ordered pairs of true positive vs. false positive responses for any desired level of observer confidence [5]. The data was analyzed using the CORROC program developed by Metz and his collaborators for correlated ROC data [6].

Observer performance was computed separately for each site in the image and for each simulated lesion. The areas, A_z, under the ROC curves and their standard deviations were calculated and used to compare the two modalities for a given site and lesion; the modalities were assumed to have no difference in their ability to depict luminance contrast if the difference in the areas of the two ROC curves was less than a preset criterion. The results were evaluated for two discrimination criteria, differences of 1.5 and 2.5 standard deviations in the areas under the ROC curves for the two contrast enhancement methods. In the following sections, more detail is given regarding the preparation of the test data and the experimental protocol.

3.2 Preparation of the Test Data

3.2.1 Selection of Normal Images

The set of normal chest images were chosen from CT scans of five separate patients who were classified as having no pathology in the areas of interest. The images were obtained on a Technicare 2060 CT scanner; the preliminary selection of images was done by an experienced radiologist. The images were taken from the scanner in digital form. The intensities in each image were calibrated in Hounsfield units, with the CT intensities approximately in the range -1000 to $+1200$; their spatial resolution was 512×512 pixels. From about 100 slices in five patients, 32 slices were chosen. Adjacent slices were avoided to maximize anatomical differences between slices and reduce the possibility of memorization of normal variation by the observer. Of the 32 slices, 8 were chosen for use as a training set, while the remaining 24 were used as the base images for the actual experiment.

3.2.2 Lesion Site Selection

In each image, four sites, two sites in the lungs and two in the mediastinum, were chosen for the insertion of artificial lesions. The criteria for site selection were the presence of appropriate natural anatomy and the prevalence of real lesions at that site in clinical practice. Similar, but not identical, sites were chosen in each base image. This selection scheme insured that the observer's task was approximately the same at each lesion site across the base set of images, a circumstance which enhances the probability that the

rating scale data can be pooled across observers and images for each site. Selection of the sites was done with the collaboration of an experienced radiologist. An example of an image with an artificial lesion inserted at one of the chosen sites is shown in Figure 4; the intensity of the inserted lesion has been exaggerated for illustrative purposes.

The selection of multiple sites within the same image in two different image fields allows the comparison of a single AHE image with multiple different windowed images; this is important in that one of the apparent diagnostic advantages of AHE is that it allows the replacement of multiple windowed images with a single view which shows diagnostic information in many fields.

3.2.3 Generation of Artificial Lesions

The characteristics of the artificial lesions inserted into the chosen sites are given in Table 1. A Gaussian intensity profile was chosen to approximate that which would be generated by a spherical tumor; the widths of the artificial lesions were chosen as appropriate to appear in the given field (lung or mediastinum).

The intensity profile for a given lesion is given by

$$I(x,y) = A(\Omega)\exp\{-(x^2+y^2)/2\sigma^2\}. \tag{1}$$

The variance of the Gaussian, σ^2, determines the width of the lesion in pixels. There are two variances given in the table for each field (lung or mediastinum); the lesions corresponding to these two variances will be referred to as the narrow and wide lesions. These widths correspond visually to a noticeable difference in lesion size. The intensity of the lesion is given by $A(\Omega)$; it depends on the local neighborhood Ω of the lesion site.

The function $A(\Omega)$ must be chosen such that the lesion intensity is not too small, in which case insufficient information will be conveyed to the observer and the observer's decisions will proceed by guesswork, nor too large, in which case the observer will make a correct choice on every trial and no discrimination of the contrast enhancement methods will be possible. A fixed lesion intensity is unacceptable, since it is well known that the ability of the eye to detect contrast is strongly dependent on both the mean intensity and the presence of structure in the near background of the lesion site. In addition, there is

evidence [7] that the far background of the image also exerts influence on the detectability of the artificial lesion.

Preliminary tests performed on the images used in this experiment indicated that the appropriate lesion intensities must be chosen on an image-by-image basis. This choice was based on a simple measure of the structural complexity, the average value of the Laplacian of the image intensity in a neighborhood Ω of the lesion site. It was assumed that the desired height, $A(\Omega)$, was given by

$$A(\Omega) = a\overline{|\nabla^2 I(x,y)|} + b, \tag{2}$$

a linear function of the average Laplacian value over the region Ω which is a square 2σ on a side centered at the position (x, y) of the lesion site.

To determine the appropriate parameters for this function, a small number of the images (4) were examined by hand; for each of these an informal method of limits experiment was performed to determine the minimum detectable lesion intensity at each site. Since the detectability of a given lesion will be affected by the processing (windowing or AHE) done on the image at the time of observation, it was necessary to perform the method of limits determination after processing of the image.

A plot was then made of the lesion intensities determined by this method against the average Laplacian. A straight line fit of the data was made and the resulting line equation was used to determine the parameters a and b. The predicted intensities were checked during the execution of training runs to verify their appropriateness. Separate equations were used for the lung and mediastinal fields; no significant difference between the wide and narrow lesions was discernible, and the same coefficients were used for both.

Two different possible lesion intensities, referred to as the bright and dim lesions, were chosen at each site. These lesions were respectively 85% and 115% of the intensity predicted by the average Laplacian measure. Thus, there were four possible lesions which could be inserted at a given lesion site, corresponding to two choices for the width and two choices for the intensity. In this experiment, only three of these possible lesions were used; it can be seen in Table 1 that the bright, wide lesion was omitted.

3.2.4 Preparation of the Trial Images

The possible characteristics of a given trial image are shown in Table 2. Each image is described by a 6-tuple of parameters that gives the base image, the lesion site, the type of lesion, lesion presence, *etc.* Thus, there are 1152 possible trial images; all these were included in the trial set. In addition, approximately 10% (128) of the possible images were chosen at random and included twice, giving a total of 1280 images in the complete set. Hence, some images were seen twice by each observer, giving a check on the consistency of a given observer in rating the images. The 6-tuples for the 1280 images were generated and used to prepare the trial images in random order; this guaranteed no discernible pattern in the order of presentation. All calculations were made in the original range of the CT data and the final images were then scaled into the display range of the display device, (0–255).

For each image, the inserted lesion was prepared by prescaling it to correspond to the local image structure as previously described. The prescaled lesion was then added (or not) to the image at the correct site. No attempt was made to simulate the disarrangement of normal structure which would be present with real clinical lesions; the linear extent of the artificial lesions was sufficiently small that no significant effects would be expected. The same lesion was inset into the upper left corner of the image to serve as a reference for the observer. This arrangement is shown in Figure 4.

In addition to the trial set described above, the 8 test images were used to generate 4 runs of 64 images each for use in training the observers. One of the 4 runs was used exclusively to familiarize the observers with the equipment and experimental procedure; the other three formed an exhaustive set of three test images over all six parameters. These test runs were used to evaluate the choice of lesion parameters (shape and intensity) to insure that they were reasonable. It was found that the scaling based on the average Laplacian produced acceptable results.

3.3 Selection of Observers

Three observers were used in this study, all board-certified physicians and experienced in the reading of chest CT scans. The observers were guaranteed that no results of the experiment would be associated with a particular observer. The experiment was designed

to occupy no more than 25 hours of reading time, including the necessary training time. The observers were trained in the use of the experimental programs and allowed to do test runs to familiarize themselves with the appearance of AHE images.

3.4 Experimental Apparatus

The trial images were displayed in grey-scale on a Tektronix 690SR RGB monitor using a Comtal 10/24 frame buffer. A VAX11/730 running the Berkeley 4.2BSD version of the Unix[1] operating system served as the host computer to control the presentation of the images and collect the data. The Tektronix monitor has very stable performance; our experience has shown that its intensity display characteristics are stable on a time scale of weeks. The monitor was calibrated and converged prior to the beginning of the experiment and checked periodically throughout the course of the experiment. The display scale was linearized using the procedures described by Johnston *et.al.* [8]. The luminance range of the monitor was 6×10^{-4} foot-lamberts for a driving intensity of 0 display units and 26.2 foot-lamberts for a driving intensity of 255; the size of the displayable screen area was 26×26 cm^2.

The Comtal 10/24 is able to display images of 512×512 pixels and contains sufficient image memory for twelve 512×512 images; it includes function memories for intensity table lookup and pipelined processors which were used to implement the linearization of the displayed image and the windowing specified by the observers. The VAX computer was equipped with a data tablet to allow the observers to supply input to the program while they were observing. Software was developed to coordinate the experiment and collect the data during the trials.

3.5 Experimental Layout

The observers were seated before the Tektronix monitor with the data tablet on a table in front of them. A light box was placed on top of the monitor facing away from the observers to provide an ambient light intensity of 3.3 lux, similar to that used for the linearization of the monitor. A video display terminal (VDT) was provided for interaction

[1] Unix is a trademark of AT&T Bell Laboratories

with the host operating system; however, once the trial run was underway, the VDT was not used. Instead, all interaction was performed using the data tablet. The observers were allowed to position themselves comfortably; no attempt was made to constrain their movements.

The physical environment of the observers was controlled insofar as this was possible in the experimental area. The room lights were extinguished and all extraneous sources of light shielded except for the light box used for ambient illumination. The observers were seated at an average distance from the screen of 1m, with the displayed image subtending an angle of 15°. The observers were required to acclimate themselves to the environment for a period of one to two minutes before beginning the experiment.

3.6 Observer Procedure

The experiment was divided into 20 sessions of length approximately one hour each. At each session, the observer was asked to rate 64 images. Before embarking upon these sessions, each observer performed 4 one hour runs for training purposes. Prior to the beginning of their training, the observers were required to read a document explaining the purpose and methods of the experiment and detailing the criteria on which their responses were to be based. After familiarization with the interaction devices and the use of the software, the observers did one training run of 64 images in which they received feedback on the correctness of their answers. This allowed them to calibrate their perceptions against the appearance of the displayed images. Following this, they performed three training runs without feedback to provide them with further experience and to allow the calibration of the experimental procedure against their performance. The results of these training runs allowed minor adjustments in the experimental procedure and the techniques for preparation of the trial runs to be made before the beginning of actual data collection.

The observers were shown each image in the trial set and asked to provide a rating of their confidence that the simulated lesion was present. Table 3 shows the definition of the five confidence categories. It should be noted that the rating scale is not symmetric; there is no response in the category "Lesion definitely present", nor is there any equivocal category. The observer was required to make a decision as to which belief was stronger,

the presence or absence of the lesion. The observers were urged to employ all five rating categories, using a 0 response on the occasions when they were most certain the lesion was not present and a 4 response when they were most certain that the lesion was present.

3.7 Analysis of Results

The result of the current experiments was a set of 1280 ratings for each observer; these can be considered as the outcome of 24 simultaneous experiments (2 fields $\times$ 2 sites per field $\times$ 3 lesions per site $\times$ 2 enhancement modalities). Each such experiment was analyzed separately using a program written by Metz *et.al.* [6], CORROC. CORROC uses the maximum likelihood estimation technique of Dorfman and Alf [9] to estimate the ROC curve due to a given set of data in the case where the images are correlated. In the present experiments, the images are correlated, since each image is shown at least twice, once as a windowed image and once as an AHE'd image. For experiments which produced similar statistical results, such as trials which used the same field and site but different lesions, the results were pooled and reanalyzed. Goodness of fit estimations were also produced, allowing an analysis of statistical confidence.

The principal statistic used in comparing two similar experiments using different contrast enhancement modalities was the area under the ROC curve, A_z; this is equal to the percentage of correct responses one would obtain from a two-alternative forced-choice experiment and gives an overall measure of the performance of a given methodology. This approach precludes the comparison of the two enhancement methodologies at different levels of observer confidence; however, the integrated area is more stable against statistical fluctuations in the data. With the relatively small sample size used in the current experiment (approximately 50 data points per experiment), this is an important criterion.

4 Results

A summary of the results for each observer is shown in Table 4; the full results are given in Tables 5–7. In these tables, the headings of each column describe the parameters of a given experiment (lesion type, site and processing modality), the integrated area under

the resulting ROC curve, and the standard deviation of the area. The standard deviations shown have been corrected for the correlation of the data by the CORROC program. There is no data pooling in these results. The sixth column shows the number of standard deviations, n_σ, by which the areas of the two ROC curves differ. The seventh column is the two-tailed p-value; it represents the confidence with which the null hypothesis, in this case that the two ROC curves have the same area, can be rejected. A small value of p indicates that the two areas being compared are unlikely to have arisen from the same underlying distribution. The final column indicates which, if either, method was found preferable. Those studies marked with a "W" showed a preference for windowing with a difference in standard deviation of at least 1.5; those marked with an "A" showed a preference for AHE at the same level. The results were evaluated at two levels; if the entry has no asterisk, the areas under the ROC curve differ by more than 1.5 standard deviations but less than 2.5 standard deviations; if the asterisk is present, the areas differ by more than 2.5 standard deviations. To allow ready comparison, the data is arranged so that experiments with similar parameters (field, lesion type, lesion site, *et.cet.*) are adjacent with only the contrast enhancement modality different. The values of 1.5 and 2.5 standard deviations correspond to two-tailed p values of 0.1336 and 0.0124 respectively.

For three cases, the CORROC program failed to converge. In these cases, the data were analyzed as if uncorrelated using the ROCFIT program of Metz; this program performs a maximum likelihood estimation on uncorrelated data. These results are shown with an "I" in the Results column. Experience showed that the values for σ derived using this method did not differ substantially from those derived with CORROC; thus it can be asserted that none of these three results shows any significant difference between the two contrast enhancement modalities.

4.1 Unpooled Results

The results of the full experiment show that in most cases there is no significant difference in the diagnostic performance between the two modalities. For a difference of 1.5 standard deviations in the lung field, windowing was superior in two experiments, AHE in two, and no significant difference was seen in 14 experiments. For 2.5 standard deviations

in the lung field, two experiments showed windowing preferable; the rest showed no significant difference. For 1.5 standard deviations in the mediastinum, windowing was prefered in six experiments, AHE in none, and no difference was seen in 11 cases. For 2.5 standard deviations in the mediastinum, three cases using windowing showed a significant difference and the remaining 13 showed no difference. Windowing did better in more cases in the mediastinum than in the lungs. This may be attributable to two factors: first, the overall data range in the mediastinum is usually less than in the lungs; hence, the amount of contrast enhancement which is achievable by windowing is greater. Second, the narrowness of the window width in the mediastinum is limited by CT noise, whereas in the lung it is limited by the presence of normal structure. Little difference is seen between observers compared to the intra-observer variation.

4.2 Results Pooled Across Lesion Sites

Given the full results, it appeared to be valid to pool the data for the same lesion when shown in the same field (lungs or mediastinum) but at different sites in the field. These results are summarized in Table 8 and bear out the results from the full data in that in the lung field for 1.5 standard deviations, no methodology seemed preferable (one experiment favored windowing, one AHE, and seven showed no significant difference) while in the mediastinum for 1.5 standard deviations five experiments showed windowing to be preferable while four showed no significant difference. For 2.5 standard deviations, only one experiment in the mediastinum showed windowing to be superior.

One difference which was easier to see in the pooled data is the difference in detectability of the different lesions. Lesions 0 and 2 appeared to be about equally detectable regardless of image field, while lesion 1 seemed to be more obvious than the other two. Lesion 1, as described in Table 1, had the greatest intensity. The difference in width of lesions 0 and 2 seemed to make no significant difference.

4.3 Time Series Analysis

One question of interest was whether the performance of the observers improved over time while conducting the experiment. It might be expected that as observer familiarity

with the AHE methodology increased, performance might improve. In order to investigate this possibility, the data pooled across the sites was arranged in the order in which the observer performed the runs; the rearranged data was divided into quarters and reanalysed. The ROCFIT program was used since the division of the data precluded correlated analysis; this was not a drawback, since the purpose was to show trends in the data rather than absolute performance. There was no apparent increase in performance for the later runs. That is, the 4 hours of training seems to have been adequate for the observers to learn fully to judge AHE'd images.

5 Discussion and Conclusions

In interpreting the results above, two circumstances must be kept in mind. First, the task which the observers were asked to perform was extremely difficult. The artificial lesions used had to be sufficiently subtle to allow some ambiguity in the interpretation of the image even when the exact location of the potential lesion was known exactly. Such lesions would be essentially undetectable in normal clinical practice. Given the preceding results, there is little question that both AHE and windowing are able to adequately depict supra-threshold contrast.

Second, in evaluating the levels of significant difference shown in Tables 5–8, it must be realized that the criterion of a difference at the 1.5σ level is very weak, with the corresponding p-value indicating a better than 1 in 8 chance of such a result arising from chance. A difference at the 2.5σ level with p-value of 0.0124 is a much stronger criterion; however, the pooled data of Table 8 show that only one of the 18 results is significant at this level. Further, an examination of the actual percent correct (column 4 of Tables 5–8) shows that in only one of the unpooled results and none of the pooled results does the probability of a correct answer differ by more than 0.2 between AHE and windowing.

Thus, the results shown indicate that there is very little difference in the ability of global linear min-max windowing and AHE to depict grey-scale contrast in an image. It seems probable that windowing holds some advantage in areas of the body (such as the mediastinum or liver) where the amount of contrast enhancement is limited by the image

noise rather than the presence of normal structure; this advantage is unlikely to be realized in clinical practice, where the windows are usually not chosen as narrowly as in the current experiments. Given the fact that AHE is automatic and reproducible, further investigation through the use of controlled clinical trials seems desirable.

5.1 Future Work

Recently, Pizer *et.al.* [10] have proposed a variant of AHE which limits the amount of contrast enhancement in areas of the image with low variability. This *Contrast-Limited Adaptive Histogram Equalization* (CLAHE) produces images in which the noise content of an image (such as is seen in the background of Figure 3) is not excessively enhanced, but in which sufficient contrast enhancement is provided for the visualization of structures within the image. Images processed with CLAHE have a more "natural" appearance and facilitate the comparison of different areas of the image. However, the reduced contrast enhancement of CLAHE may hinder the ability of an observer to detect the presence of some significant grey-scale contrast. Psychophysical investigations similar to those described in this paper are currently underway to determine the relative ability of AHE and CLAHE to depict grey-scale contrast.

Given the promise which AHE has shown as an effective, automatic method for contrast enhancement, a number of projects have been been undertaken to implement AHE on special-purpose hardware. An implementation of AHE using Very Large Scale Integrated (VLSI) circuits has been proposed which will process a large image in under one second [11].

6 Acknowledgements

The authors gratefully acknowledge the following for useful discussions and critiques of an earlier version of this paper: R. Eugene Johnston, Frederick P. Brooks, Jr., and Jan J. Koenderink. We thank Donald P. Schreiter for photographic assistance.

7 References

[1] S. M. Pizer, J. B. Zimmerman, and E. V. Staab, "Adaptive Grey Level Assignment in CT Scan Display," *Journal of Computer Assisted Tomography*, vol. 8, pp. 300–308, 1984.

[2] D. J. Ketcham, R. Lowe, and W. Weber, "Real-Time Enhancement Techniques," *Seminar on Image Processing*, Hughes Aircraft, pp. 1–6, 1976.

[3] R. Hummel "Image Enhancement by Histogram Transformation," *Computer Graphics and Image Processing*, vol. 6, pp. 184–195, 1977.

[4] S. M. Pizer "An Automatic Intensity Mapping for the Display of CT Scans and Other Images," *Proc. VIIth International Meeting on Information Processing in Medical Imaging*, pp. 276–309, 1983.

[5] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, New York: Robert E. Krieger, 1974.

[6] C. E. Metz, P. Wang, and H. B. Kronman, "A New Approach for Testing the Significance of Differences Between ROC Curves Measured from Correlated Data," *VIIIth Conference on Information Processing in Medical Imaging*, F. Deconinck (ed.), The Hague: Martinus Nijhoff, 1984.

[7] G. J. Van der Wildt and R. G. Waarts, "Contrast Detection and Its Dependence on the Presence of Edges and Lines in the Stimulus Field," *Vision Research*, vol. 23, Pergamon Press, Ltd, pp. 821–830, 1983.

[8] R. E. Johnston, J. B. Zimmerman, D. C. Rogers, and S. M. Pizer, "Perceptual Standardization," *Third International Conference on Picture Archiving and Communications Systems (PACS III) for Medical Applications*, vol. 536, SPIE, 1985.

[9] D. D. Dorfman and E. A. Jr., "Maximum-Likelihood Estimation of Parameters of Signal-Detection Theory and Determination of Confidence Intervals - Rating Method Data," *Journal of Mathematical Psychology*, vol. 6, pp. 487–496, 1969.

[10] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive Histogram

Equalization and its Variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, pp. 355–368, 1987.

[11] S. M. Pizer, J. D. Austin, J. R. Perry, and J. B. Zimmerman, "Adaptive Histogram Equalization for Automatic Contrast Enhancement of Medical Images," *Fourth International Conference on Picture Archiving and Communications Systems (PACS IV) for Medical Applications*, vol. 626, SPIE, 1986.

List of Captions

Figure 1: Global linear min-max windowing.

Figure 2: Adaptive Histogram Equalization. The contextual region shown is a square of $n \times n$ pixels about a pixel at location (x, y).

Figure 3: Adaptive Histogram Equalization. The original image is shown in the upper left. The upper right image has been processed with AHE. The lower left image is windowed to show the lungs, the lower right to show the mediastinum.

Figure 4: Artificial lesion inserted into image of Figure 3. The site shown is typical; the intensity of the lesion has been exaggerated for illustrative purposes.

Table 1: Parameters for the Artificial Lesions. Widths are given in pixels. The height factor multiplies the required intensity predicted by the average value of the Laplacian (see text).

Table 2: Parameters for the Trial Images.

Table 3: Rating Criteria for the ROC Response Categories.

Table 4: Summary of results for the unpooled rating scale data. Each column gives the preferred enhancement method, windowing (W) or AHE (A), at the shown discrimination level. For each field, lungs or mediastinum, there were six results, corresponding to 2 sites within the field $\times$ 3 possible artificial lesions.

Table 5: Results for Observer 1. The columns indicate the lesion type, site, and processing method; the area under the ROC curve and its standard deviation; the number of corrected standard deviations by which the areas differ; the two-tailed p value; and the result. In the results column, the letter indicates the preferred method (W = windowing, A = AHE) at the 1.5σ level; the addition of an asterisk indicates that the results were significant at the 2.5σ level. An I in the results column indiciates that the data was analyzed as if uncorrelated. The lesion types are as given in Table 1. No data pooling was performed.

Table 6: Results for Observer 2. No data pooling.

Table 7: Results for Observer 3. No data pooling.

Table 8: Summary of results after pooling of data across lesions sites. The columns are as given in Table 4. Three results are given, corresponding to the three different lesion types as given in Table 1.

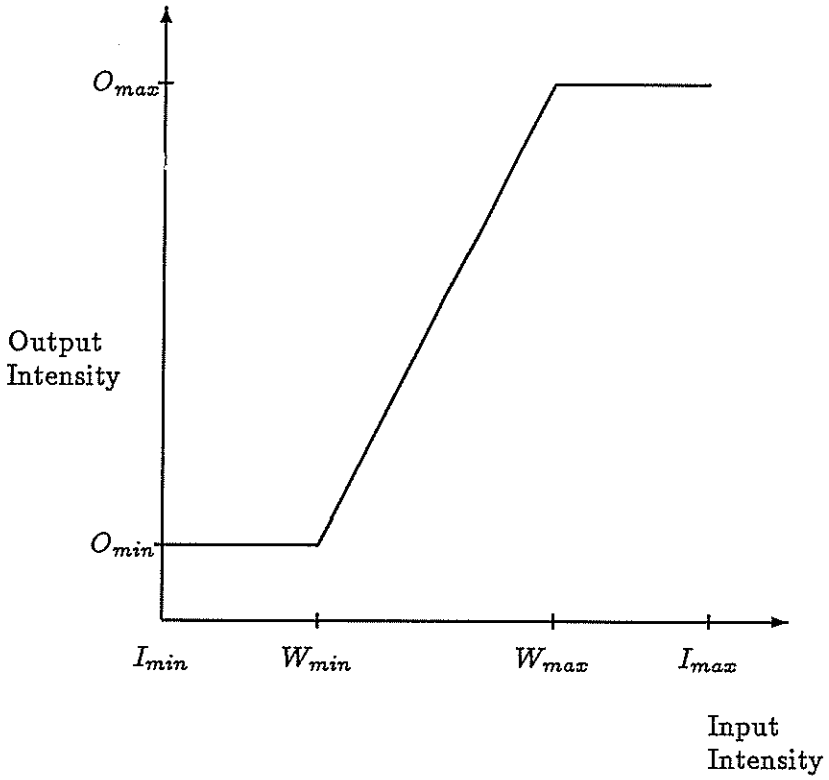

Figure 1: Global linear min-max windowing.

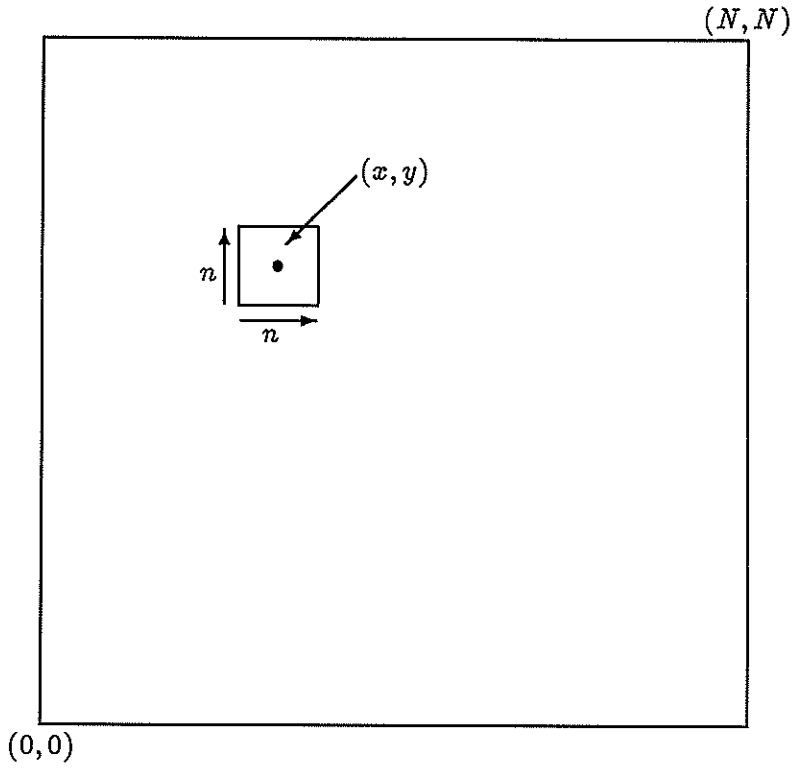

Figure 2: Adaptive Histogram Equalization. The contextual region shown is a square of $n \times n$ pixels about a pixel at location (x, y).

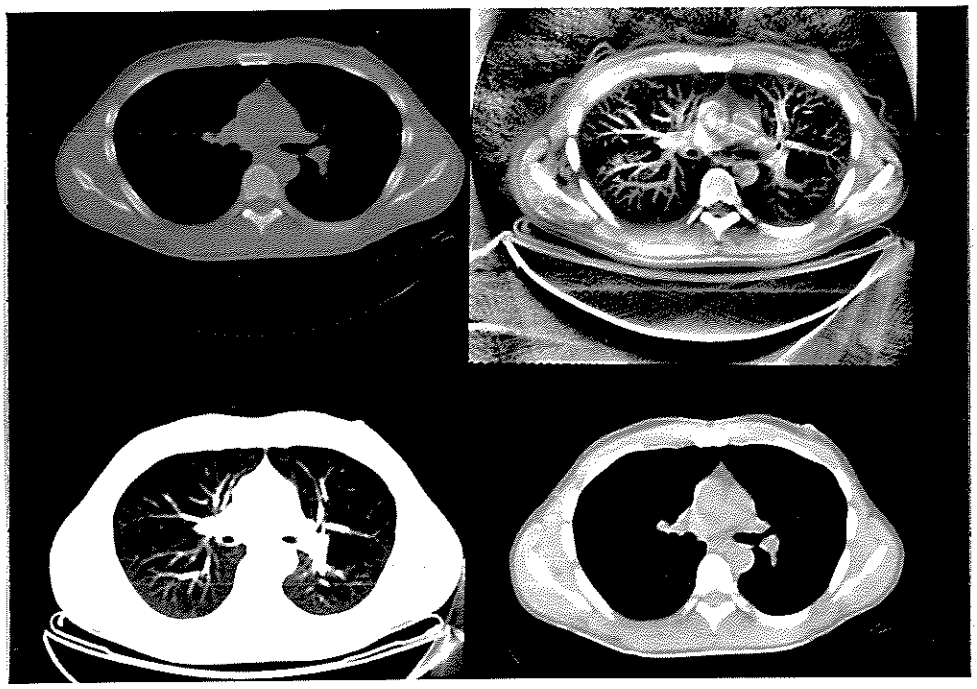

Figure 3: Adaptive Histogram Equalization. The original image is shown in the upper left. The upper right image has been processed with AHE. The lower left image is windowed to show the lungs, the lower right to show the mediastinum.

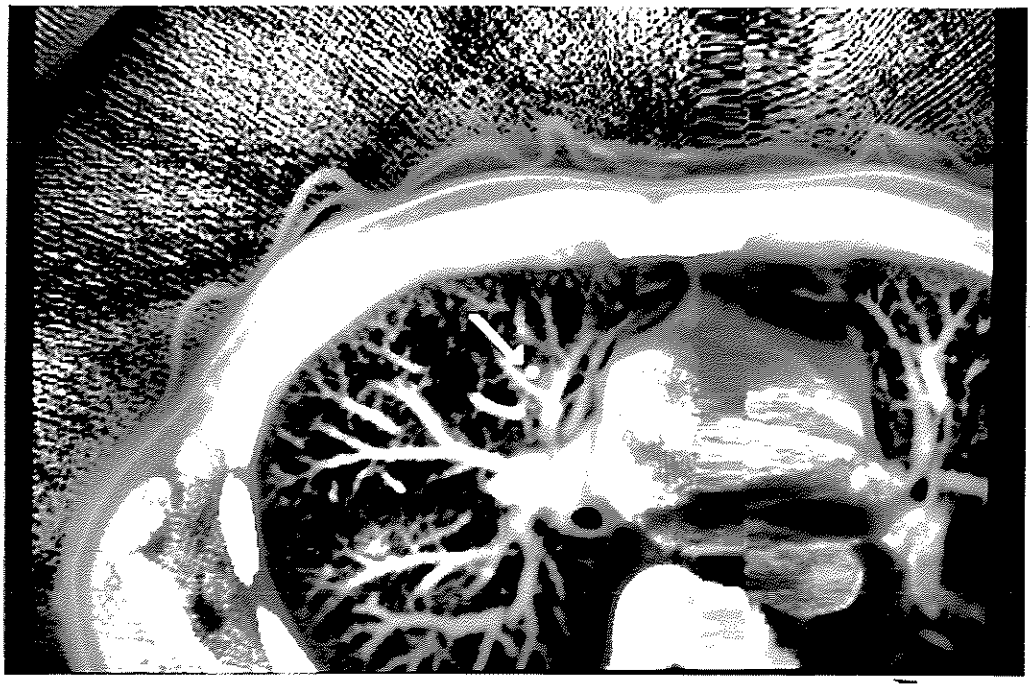

Figure 4: Artificial lesion inserted into image of Figure 3. The site shown is typical; the intensity of the lesion has been exaggerated for illustrative purposes.

Lungs		
Lesion Number	Width (σ)	Height factor
0	1.41	0.85
1	2.0	1.15
2	2.0	0.85

Mediastinum		
Lesion Number	Width (σ)	Height factor
0	2.0	0.85
1	2.82	1.15
2	2.82	0.85

Table 1: Parameters for the Artificial Lesions. Widths are given in pixels. The height factor multiplies the required intensity predicted by the average value of the Laplacian (see text).

Lesion Parameter	Values	Comments
Base Image	24	Chest CT scans
Lesion Field	2	Lungs or Mediastinum
Lesion Type	3	As given in Table 1
Lesion Site	2	Site within the field
Processing	2	AHE or windowing
Lesion Presence	2	Present or absent
Redundancy Factor	10%	Number of repeated images
Total	1280	All combinations

Table 2: Parameters for the Trial Images

Rating	Confidence
0	Definitely not present
1	Probably not present
2	Possibly not present
3	Possibly present
4	Probably present

Table 3: Rating Criteria for the ROC Response Categories

Observer 1					
Field	W (1.5σ)	W (2.5σ)	A (1.5σ)	A (2.5σ)	None
Lungs	0	0	1	0	5
Mediastinum	0	2	0	0	4

Observer 2					
Field	W (1.5σ)	W (2.5σ)	A (1.5σ)	A (2.5σ)	None
Lungs	0	1	0	0	5
Mediastinum	1	1	0	0	4

Observer 3					
Field	W (1.5σ)	W (2.5σ)	A (1.5σ)	A (2.5σ)	None
Lungs	0	1	1	0	4
Mediastinum	2	0	0	0	4

Table 4: Summary of results for the unpooled rating scale data. Each column gives the preferred enhancement method, windowing (W) or AHE (A), at the shown discrimination level. For each field, lungs or mediastinum, there were six results, corresponding to 2 sites within the field $\times$ 3 possible artificial lesions.

Observer 1

Lungs

Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.7352	0.0694	0.6380	0.5235	
0	0	A	0.6808	0.0764	0.6380	0.5235	
1	0	W	0.7877	0.0628	-0.1635	0.8701	
1	0	A	0.7980	0.0650	-0.1635	0.8701	
2	0	W	0.6478	0.0747	-0.5953	0.5516	
2	0	A	0.7038	0.0741	-0.5953	0.5516	
0	1	W	0.7152	0.0680	-1.5024	0.1330	
0	1	A	0.8303	0.0633	-1.5024	0.1330	A
1	1	W	0.8866	0.0472			I
1	1	A	0.8836	0.0512			
2	1	W	0.7478	0.0682	-0.2562	0.7978	
2	1	A	0.7694	0.0672	-0.2562	0.7978	

Mediastinum

Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.8567	0.0513	2.5621	0.0104	W*
0	0	A	0.6619	0.0761	2.5621	0.0104	
1	0	W	0.8205	0.0652			I
1	0	A	0.7499	0.0735			
2	0	W	0.6616	0.0748	0.5819	0.5606	
2	0	A	0.6031	0.0893	0.5819	0.5606	
0	1	W	0.7293	0.0705	0.6991	0.4845	
0	1	A	0.6720	0.0748	0.6991	0.4845	
1	1	W	0.7526	0.0693	2.5052	0.0122	W*
1	1	A	0.5795	0.0797	2.5052	0.0122	
2	1	W	0.6703	0.0769	1.0135	0.3108	
2	1	A	0.5993	0.0791	1.0135	0.3108	

Table 5: Results for Observer 1. The columns indicate the lesion type, site, and processing method; the area under the ROC curve and its standard deviation; the number of corrected standard deviations by which the areas differ; the two-tailed p value; and the result. In the results column, the letter indicates the preferred method (W = windowing, A = AHE) at the 1.5σ level; the addition of an asterisk indicates that the results were significant at the 2.5σ level. An I in the results column indiciates that the data was analyzed as if uncorrelated. The lesion types are as given in Table 1. No data pooling was performed.

Observer 2

Lungs

Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.6865	0.0736	-0.4104	0.6815	
0	0	A	0.7237	0.0720	-0.4104	0.6815	
1	0	W	0.8679	0.0465	0.1426	0.8866	
1	0	A	0.8609	0.0496	0.1426	0.8866	
2	0	W	0.8506	0.0524	2.6395	0.0083	W*
2	0	A	0.6378	0.0728	2.6395	0.0083	
0	1	W	0.7973	0.0573	-0.7875	0.4310	
0	1	A	0.8440	0.0516	-0.7875	0.4310	
1	1	W	0.9371	0.0318	1.3777	0.1683	
1	1	A	0.8779	0.0456	1.3777	0.1683	
2	1	W	0.7558	0.0708			I
2	1	A	0.8268	0.0624			

Mediastinum

Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.9061	0.0503	1.9669	0.0492	W
0	0	A	0.7840	0.0609	1.9669	0.0492	
1	0	W	0.8929	0.0430	2.5087	0.0121	W*
1	0	A	0.7323	0.0693	2.5087	0.0121	
2	0	W	0.6924	0.0713	-0.1845	0.8536	
2	0	A	0.7067	0.0709	-0.1845	0.8536	
0	1	W	0.7587	0.0650	0.4053	0.6852	
0	1	A	0.7249	0.0704	0.4053	0.6852	
1	1	W	0.7698	0.0632	0.3883	0.6978	
1	1	A	0.7321	0.0704	0.3883	0.6978	
2	1	W	0.6363	0.0754	-0.6317	0.5276	
2	1	A	0.6881	0.0744	-0.6317	0.5276	

Table 6: Results for Observer 2. No data pooling.

Observer 3

Lungs

Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.7827	0.0640	2.5205	0.0117	W*
0	0	A	0.5963	0.0789	2.5205	0.0117	
1	0	W	0.6363	0.0960	-0.7300	0.4654	
1	0	A	0.7088	0.0683	-0.7300	0.4654	
2	0	W	0.5517	0.0837	-1.0492	0.2941	
2	0	A	0.6380	0.0737	-1.0492	0.2941	
0	1	W	0.7760	0.0897	-0.3544	0.7230	
0	1	A	0.8074	0.0577	-0.3544	0.7230	
1	1	W	0.6269	0.0955	-1.7126	0.0868	
1	1	A	0.7915	0.0597	-1.7126	0.0868	A
2	1	W	0.8184	0.0782	0.3866	0.6990	
2	1	A	0.7851	0.0619	0.3866	0.6990	

Mediastinum

Type	Site	Proc	Area	σ	n_σ	pval	Result
0	0	W	0.7544	0.0635	0.9862	0.3241	
0	0	A	0.6807	0.0705	0.9862	0.3241	
1	0	W	0.8094	0.0618	1.2161	0.2239	
1	0	A	0.7258	0.0685	1.2161	0.2239	
2	0	W	0.6595	0.0733	-0.1898	0.8495	
2	0	A	0.6768	0.0812	-0.1898	0.8495	
0	1	W	0.8078	0.0611	2.2891	0.0221	W
0	1	A	0.6362	0.0766	2.2891	0.0221	
1	1	W	0.6350	0.0771	1.6268	0.1038	W
1	1	A	0.5249	0.0830	1.6268	0.1038	
2	1	W	0.6073	0.0767	-0.3827	0.7020	
2	1	A	0.6405	0.0771	-0.3827	0.7020	

Table 7: Results for Observer 3. No data pooling.

Observer 1					
Field	W (1.5σ)	W (2.5σ)	A (1.5σ)	A (2.5σ)	None
Lungs	0	0	0	0	3
Mediastinum	1	1	0	0	1

Observer 2					
Field	W (1.5σ)	W (2.5σ)	A (1.5σ)	A (2.5σ)	None
Lungs	0	0	0	0	3
Mediastinum	1	0	0	0	2

Observer 3					
Field	W (1.5σ)	W (2.5σ)	A (1.5σ)	A (2.5σ)	None
Lungs	1	0	1	0	1
Mediastinum	2	0	0	0	1

Table 8: Summary of results after pooling of data across lesions sites. The columns are as given in Table 4. Three results are given, corresponding to the three different lesion types as given in Table 1.