Comparison of clinical findings between intensity-windowed versus CLAHE presentation of chest CT images

BM Hemminger, RE Johnston, K Muller*, D Taylor*,
M Mauro, M Schiebler, E Pisano

University of North Carolina at Chapel Hill
Departments of Radiology and Biostatistics*
Chapel Hill, NC, 27599-7510

ABSTRACT

We are investigating how radiologist's readings of standard intensity windowed (IW) chest Computed Tomography (CT) films compare with readings of the same images processed with Contrast Limited Adaptive Histogram Equalization (CLAHE). Previously reported studies where CLAHE has been tested have involved detection of computer generated targets in medical images. Our study is designed to evaluate CLAHE when applied to clinical material and to compare the diagnostic information perceived by the radiologists from CLAHE processed images to that from the conventional IW images.

Our initial experiment with two radiologists did not yield conclusive results, due in part, to inadequate observer training prior to the experiment. The initial experimental protocol was redesigned to include more in-depth training. Three new radiologist observers were recruited for the follow-up study. Results from the initial study are reviewed and the follow-up study is presented. In the new study we find that while CLAHE and IW are not statistically significantly different overall, there are specific clinical findings where the radiologists were less comfortable reading CLAHE presentations. Advantages and disadvantages of using CLAHE as a replacement or as an adjunct to IW are discussed.

1. INTRODUCTION

A family of image enhancement techniques have been developed at the University of North Carolina (UNC), beginning with Adaptive Histogram Equalization (AHE) originated by Steve Pizer[1]. After preliminary clinical testing and evaluation, AHE was extended to Contrast Limited Adaptive Histogram Equalization (CLAHE)[2,3] to allow adjustment of the amount of enhancement. Limiting the enhancement reduced the presentation of noise and reconstruction artifacts that were disconcerting to the physician. Clinicians were then able to control the enhancement process allowing significant enhancement while maintaining presentations similar to what they are accustomed. Also the CLAHE algorithm was implemented in hardware to decrease the execution time from two hours on current workstations to 4 seconds.

Early work at UNC showed that AHE was equivalent to IW for the detection of computer generated lesions inserted into the lung area of chest CT images[4]. Other work at Arizona has shown similar results for chest Xrays[5]. A comprehensive study in Radiation Oncology at the UNC using real clinical material was just completed where CLAHE portal images were compared to the traditional port films. The study concluded that CLAHE processed port films were better than unprocessed films[6]. However, studies have not been carried out to evaluate whether CLAHE could be used effectively in the clinic to replace intensity windowing or to serve as an adjunct to it. This paper describes our results in evaluating whether CLAHE is effective as a replacement or as an adjunct to IW for the presentation of chest CT images.

At our institution, chest CT images are normally presented on a standard light box or alternator with 20 to 40 slices, and two to three different intensity window presentations (soft-tissue, lung, liver). One advantage we might hope to accrue from using CLAHE as the presentation method is quicker readings because of the

display of detail in all areas of the chest at once (soft-tissue, lung, liver, bone, etc). Efficiency should come from viewing a single set of CLAHE'd images, and not having to handle and view the many separate IW sets [figure 1]. Similarly, it may allow better comparison of the same anatomy in a single image. A second advantage might be the ability to see things not detected with IW. CLAHE shows information not presented by IW because IW clips the original information and displays only a range of the original values. If these two advantages could be shown, then we might be able to provide faster and more informed chest CT readings than are currently done in the clinic. This could result in higher throughput and better patient care, respectively.

Figure 1 Single image shown unprocessed (upper left), CLAHE'd (upper right),
lung IW'd (lower left), mediastinum IW'd (lower right)

We carried out an initial study to compare clinical information reported by radiologists reading IW and CLAHE chest CT studies. Radiologists read CLAHE and IW images from film. To compare the two methods, the radiologists completed clinical findings forms for each study they read. Two radiologists were involved in the initial study. However, the study was flawed because the radiologists operated at different confidence levels when reading the CLAHE'd images than they did when reading the IW images. The experimental paradigm was based on the assumption that the radiologists used the same ranking criteria for CLAHE and IW readings. The observed variation in ranking implied inadequate training of the radiologists when they read the new (CLAHE) images[7]. We redesigned the study to include a more extensive training period with experimenter interaction with the radiologists to better assess their training and to circumvent the dependency on ranking consistency. Three different radiologists were recruited for the new study.

2. METHODS

From the initial study we learned that we needed to provide more training in reading CLAHE. To provide training in CLAHE that would equal the radiologist's experience in IW would be unrealistic. Given a limited number of studies and the scarcity of radiologists observer time, we choose to (1) increase the number of training sets, (2) to improve the radiologists learning during the training cases by spending more time comparing between CLAHE and IW presentations, (3) to provide interactive feedback from the experimenter, and (4) to attempt to evaluate their comfort at reading CLAHE images at the end of the training. This training methodology is likely equivalent to that received by the radiologist in a one day continuing education seminar on a single topic.

2.1 Study design

The image data sets were separated into two groups, the training set and the trials set. The trials set was the same data previously used in the initial study. All the image data sets had been acquired on our Technicare 2060 CT scanner. Patient studies were selected at random from the workload of the clinic. The images were saved to magnetic tape, read into our research computers, processed with the CLAHE algorithm, written back out to tape in the same Technicare format, and then printed to film on a Matrix film formatter. We processed 13 new studies from the Technicare scanner to use for the training set. These were gathered in the same fashion except that we were no longer able to print the processed images to film because of hardware and software changes. The training was instead done using our FilmPlane electronic workstation[8]. This workstation has been shown be equivalent to film for reading CT images, albeit somewhat slower[9]. Both the CLAHED and IW versions of the training films were presented on FilmPlane.

One patient study from the training set was used to acquaint users with the use of the FilmPlane workstation and to go over the format of the training studies. The remaining twelve studies were used for the actual training series. Because we were considering only clinical findings in the chest, images from the abdominal area that were below the chest area were masked off (on film) or not included (on FilmPlane).

2.2 Image Preparation

The image data was processed using true CLAHE (as compared to interpolated CLAHE, a form often used because of its significantly faster computation time). The CLAHE processing was done on MAHEM, a hardware device designed and built at UNC specifically for processing medical images with CLAHE in real time. The images were processed with CLAHE parameters of a contrast limit factor of 20 and a contextual region size of 64 pixels. These parameters were selected to provide good enhancement of the images while not overemphasizing the noise or reconstruction artifacts. Although one might choose to fine tune the CLAHE parameters in the same way one fine tunes IW settings, for the most part our experience with chest CT images shows that one can find a single choice of parameters that works reasonably well for all chest CT images. In addition, we specifically wanted to test the feasibility of a single CLAHE setting because if multiple settings are required then CLAHE loses the advantage of presenting only a single set of images versus the multiple windowed sets of images used in intensity windowing. These same parameters were used in the initial study as well as for the 29 trials and 13 training cases. The values were selected by a radiologist who is a consultant for this research, but not an observer.

2.3 Training Protocol

Radiologists were first shown the CLAHE presentation of the patient study on the video screen followed by the IW presentation of the patient study on film. As the CLAHE'd images are presented they were given the *initial reading* clinical findings form. The clinical findings form is a list of 27 check off items that our radiologists mentally go through when performing a chest CT study. The clinical findings list was generated by our clinical radiologists who are experts in reading chest CT. The radiologists would score each finding on a scale of 1 to 5, with 1, 2, 3, 4 and 5 representing *normal, probably normal, possibly abnormal, probably abnormal* and *abnormal*, respectively. They recorded their impressions of the study on the initial reading findings form just as if they were reading it in the clinic.

After they had completed the clinical findings form for the CLAHE presentation, the same patient study was additionally presented in IW format and the radiologist was given the *second reading* findings form and asked to complete it. Their task for the second reading findings form was to review each clinical finding, given both presentations, and indicate which items they would score differently, and how each item would be scored. Additionally, we asked them to detail specific differences between the presentations and to indicate the exact scan slices involved when they scored the clinical findings differently between the first and second findings forms.

Each session involved reading several patient studies, with the amount of time spent during a single session ranging from 1/2 hour to 1 hour to avoid fatigue. Additionally, sessions were limited to two a week. We also attempted not to have long periods between sessions once the training was complete in order to minimize their loss of familiarity with the CLAHE presentation. However, because of clinical demands on the radiologists we had a break of several weeks between sessions for two of the radiologists.

In the training phase we always presented CLAHE first. This was intended to maximize their experience with making their initial readings on CLAHE alone, before incorporating IW for the second reading. This would allow us to query them at different stages of their training to assess their comfort in reading CLAHE studies. In order to improve their understanding of the CLAHE presentation we encouraged them to study the scans of each patient study in depth while both CLAHE and IW presentations were available. The experimenter also interacted with them during the second reading to answer questions or to point out areas where CLAHE presentations were different from IW. We felt this enhanced their learning during the training sessions by encouraging them to explore the differences and by our being able to prompt them to study certain areas of difference. This prompting allowed us to ascertain how well the radiologists were understanding the CLAHE presentations, and whether there were significant differences between the radiologist's impressions.

Because we could not provide training by association with absolute truth (pathology reports, autopsy data, etc.) for this study, we provided training by association with what they were already expert in, i.e. IW presentations.

2.4 Study Protocol
The study protocol was essentially the same as the training protocol. The differences were (1) both presentations were on film; (2) the studies were randomly preselected so that the initial presentation was either CLAHE or IW; (3) there was no interaction with the experimenter; (4) while the radiologists were expected to note differences between the presentations they were not expected to spend time studying them in depth; (5) the times taken to read the first presentation and the combined presentation were recorded.

The training set consisted of 12 patient studies. The study set consisted of 29 patient studies. The study set was additionally subdivided into three groups of equal size based on the number of disease findings in each patient study. The three groups were *least findings*, *medium findings*, and *most findings*. These groupings did not affect the order of presentation, and were recorded only to allow possible grouping by case type in the statistics measures.

During the training sessions, verbal questions and comments were encouraged from the observers. This provided a large amount of feedback from the users. Additionally, at the end of each experimental session the radiologists were queried for feedback as well. Finally, at the conclusion of the last run in the study, an exit interview was conducted to ask the observers several specific questions regarding CLAHE and its usefulness from their perspective.

3. DATA ANALYSIS

We attempted to provide statistically valid measures of (1) whether CLAHE performed equivalently to IW for clinical material, (2) whether CLAHE would be a useful adjunct to IW in the clinic, and (3) whether reading from CLAHE presentations would be faster than reading from IW presentations. The clinical task of reading and dictating was modeled with the filling out of the findings list. Results were collected for each of the three radiologists for each of the 27 clinical findings over all 29 patient studies. For each patient study initial reading and second reading score sheets were completed by the radiologists. The initial score (1-5) came after the first reading, and the revised score (1-5) after the second combined reading. Results were tallied onto 5x5 grids [figure 2], with an initial score of n and a second score of m being mapped at row n, column m. There was a 5x5 grid for each clinical finding for each radiologist for each of IW or CLAHE as the initial presentation. The raw data was input into a statistical program (SAS) where we could perform reductions and extractions of the data and calculate statistical properties of it.

Second \ Initial	Normal	Probably Normal	Possibly Abnormal	Probably Abnormal	Abnormal
Normal	1,1	1,2	1,3	1,4	1,5
Probably Normal	2,1	2,2	2,3	2,4	2,5
Possibly Abnormal	3,1	3,2	3,3	3,4	3,5
Probably Abnormal	4,1	4,2	4,3	4,4	4,5
Abnormal	5,1	5,2	5,3	5,4	5,5

Figure 2   5x5 grid layout

Second \ Initial	Normal	Abnormal
Normal	1,1	1,2
Abnormal	2,1	2,2

Figure 3   2x2 grid layout

Because of the sparsity of the data in the 5x5 grids our first step was to map the 5x5 grids down to a smaller size. We grouped the normal and probably normal cells into one group, and the possibly abnormal, probably abnormal and the abnormal cells into another group. This resulted in a 2x2 grid [figure 3] with cells corresponding to the categories of *normal/normal*, *normal/abnormal*, *abnormal/normal* and *abnormal/abnormal*. All further calculations were performed using the 2x2 grids. Elements of (1,1) cell are cases where the clinical finding evaluated to normal in both the first and second presentation. Elements in (2,2) cell are cases where the clinical findings evaluated to abnormal in both the first and second presentations. Elements in the (1,2) cell are cases where the clinical finding report from the first presentation was normal, but the clinical finding report from second presentation combined with the first presentation was abnormal. Elements in the (2,1) cell are cases where the clinical finding report from the first presentation was abnormal, but the clinical finding report from the second presentation combined with the first presentation was normal.

CLAHE and IW equivalent

To test whether CLAHE could be considered as a replacement for IW in the clinic we examined whether the clinical findings scores differed from the initial reading to the second reading. Kappa statistics were used to measure the agreement between scores reported on the initial readings form (after the observer has seen only the initial presentation) and second readings form (after the observer has seen both presentations). Kappa values of approximately .75 to 1.0 may be taken to represent excellent agreement beyond chance, values below .40 or so may be taken to represent poor agreement beyond chance, and values between .40 and .75 may be taken to represent fair to good agreement beyond chance[10]. Kappa values were computed for each of the following categories:

I	each radiologist	vs.	each clinical finding
II	all radiologists	vs.	each clinical finding
III	all radiologists	vs.	groups of clinical findings
IV	each radiologists	vs.	all clinical findings
V	all radiologists	vs.	all clinical findings

The data from I and II are not listed in the paper because of length considerations. In the following tables CLAHE followed by IW presentations are referred to CLAHE/IW and IW followed by CLAHE presentations are referred to as IW/CLAHE. For each of the two presentation orderings, CLAHE/IW and IW/CLAHE, a maximum of only 15 readings were made per finding and observer. Consequently, the calculated standard errors for individual kappa statistics proved prohibitively large for confidence interval based testing. Although grouping across readers and findings (categories II - V) presents descriptive characteristics of the data with less variant kappa statistics, the resulting standard errors are conservative due to a lack of independence among findings and reader observations (since each reader was given the same set of films per finding). Thus confidence interval based testing of differences in kappa statistics was not feasible with this data.

Table 1 shows Kappa statistics for V where we have grouped over all observers and all findings. Kappa statistics for IV where we have grouped over all clinical findings for each observer are shown in table 2. In these more encompassing categories (IV and V), where all the clinical findings are included, we see that there is good to excellent agreement between the first and second readings.

Reader	Order	Kappa	Standard Error
All	CLAHE/IW	0.71	0.03
All	IW/CLAHE	0.86	0.02

Table 1 Kappa statistics for all readers for all findings

Reader	Order	Kappa	Standard Error
1	CLAHE / IW	0.64	0.05
1	IW / CLAHE	0.75	0.04
2	CLAHE / IW	0.79	0.05
2	IW / CLAHE	0.96	0.02
3	CLAHE / IW	0.71	0.05
3	IW / CLAHE	0.93	0.03

Table 2 Kappa statistics by reader over all findings

However, looking carefully at the CLAHE first versus the IW first categories in table 2 and individual findings in **I** and **II**, we find that the agreement is generally not as good for CLAHE first presentations as it is for IW first presentations. This implies that radiologists are more likely to make corrections to initial CLAHE readings after the addition of the IW presentation rather than the other way around. Table 3 contains the Kappa statistics for two groups from category **III**, lung and mediastinum. These anatomical groupings were chosen because they commonly are the areas where disease is most prevalent in chest CT clinical findings. In these areas where the incidence of disease is the highest we find that CLAHE/IW has less agreement than IW/CLAHE (table 3). Further, we find fair to poor agreement for CLAHE/IW for the lung grouping.

Area	Order	Kappa	Standard Error
Lung	**CLAHE / IW**	0.49	0.06
Lung	**IW / CLAHE**	0.73	0.04
MS	**CLAHE / IW**	0.83	0.05
MS	**IW / CLAHE**	0.91	0.05

Table 3 Kappa statistics by disease area groupings of findings over all readers

Since the cases for the study were taken at random from the clinic, a large portion of the cases were normals. The normals help ascertain when there were *false positives* (clinical findings reported in the initial reading but not in the second reading) and *false negatives* (clinical findings reported in the second reading but not the first reading). However, because a large portion of the cases did not have significant disease we find that there is good agreement overall mainly because of the contributions from the (1,1) cell of the 2x2 grid where both the initial and the second reading recorded normal. Table 4 shows the cell counts for 2x2 grid for all radiologists for all clinical findings. Thus while there appears to be agreement in the overall category (table 1), areas where more disease is present tend to show considerably less agreement (table 3).

Normal / Normal CLAHE 955 IW 912	*Normal / Abnormal* CLAHE 43 IW 31
Abnormal / Normal CLAHE 58 IW 13	*Abnormal /Abnormal* CLAHE 158 IW 175

Table 4 Cell counts for 2x2 grid for all radiologists for all findings

To further investigate whether the individual readers preferred IW/CLAHE or CLAHE/IW, an intra-reader sign test was performed. The null hypothesis was that there was no difference in agreement between IW/CLAHE and CLAHE/IW over the 27 different clinical findings. The results are shown in table 5.

Reader	favor IW/Clahe	favor Clahe/IW	ties	p-value
1	9	5	13	0.124
2	12	3	12	0.026
3	6	3	18	0.221

Table 5 Intra-reader sign test

From this we would reject the null hypothesis for reader 2 for an alpha value of 0.05, concluding that IW/CLAHE has better agreement than CLAHE/IW. While readers 1 and 3 numerically have more agreements for IW followed by CLAHE as well, statistically we would fail to reject the null hypothesis for them, possibly due to limited sample sizes.

CLAHE as an adjunct

To answer the question of whether CLAHE would be useful as an adjunct we examine the tables to see if the clinical findings differed between initial readings of IW alone and IW/CLAHE. Differences between the two would indicate that the radiologist changed their clinical report due to the addition of the CLAHE presentation. This would tell us that CLAHE may be useful as an adjunct, although we would want to follow up with areas where CLAHE changed the clinical report and correlate them with anatomical truth. We again see that overall there is good agreement between IW and IW followed by CLAHE. However, as before, this is weighted towards the normal/normal cell (1,1) as seen in table 4. In the two subgroups (table 3) where we have more disease, we again find less agreement than in the overall picture. In this case though the disagreement is less. The mediastinum disease area has excellent agreement (table 3). From the data for each radiologist versus each clinical finding, we find that the lung area has the worst agreement between IW and IW followed by CLAHE, however, from table 3 we can see that agreement for this group is essentially excellent as well. Thus while the agreement is less in the areas with more disease, we still find that for these areas as well as overall there is excellent agreement between IW and IW followed by CLAHE. From this we would conclude that CLAHE did not change the clinical findings when it was made available to the radiologist in addition to IW.

CLAHE as faster

To test whether CLAHE readings were faster than IW readings we recorded the reading times. Timing was begun when the radiologist was handed the folder and concluded when they indicated they had finished filling out the initial reading clinical findings form. In table 6 the results from a general linear model repeated measures analysis for CLAHE and IW readings are listed. For the repeated measures test, the null hypothesis was that there was no difference in reading times between CLAHE initial readings and IW initial readings. A P value of 0.0439 was obtained with an f statistic of 21.27 with one numerator and two denominator degrees of freedom. Thus we would reject the null hypothesis at the 0.05 level and conclude that the time to complete CLAHE readings differs from completing IW readings. From the mean times we see that CLAHE is faster and thus CLAHE readings are significantly faster than IW readings.

We also tested for the null hypothesis of no difference in reading time among the different rankings (least findings, medium findings, most findings). For an unadjusted f statistic of 21.65 with two numerators and four denominator degrees of freedom, the Greenhouse-Geisser adjusted P-value was 0.0269 (with epsilon equal to 0.6283). Thus we reject the null hypothesis and conclude that reading time differs with the amount of disease present (table 7).

Order	Samples	Mean	Std Dev	Minimum	Maximum
IW/CLAHE	42	154.50	72.31	63.00	375.00
CLAHE/IW	45	122.29	71.17	32.00	420.00

Table 6 Reading times related to presentation method

Rank	Samples	Mean	Std Dev	Minimum	Maximum
least	30	117.30	62.19	32.00	287.00
medium	30	141.83	71.23	61.00	375.00
most	27	156.22	82.80	60.00	420.00

Table 7 Reading times related to degree of disease present

4. DISCUSSION

4.1 Training
All the observers progressed from initial discomfort at reading CLAHE images to reasonable comfort reading CLAHE images. CLAHE images appear somewhat different from IW images and require some adjustment [figure 1]. By the conclusion of the training series each radiologist said that they could read a clinical case using CLAHE. However, it became apparent during the training that in some cases this meant they were not able to see certain areas the same way they were accustomed to viewing them via IW and settled for interpreting CLAHE presentations as all normal or all abnormal for that clinical finding while they distinguished between them on IW. Also, during the training it became clear that there were several specific areas where the radiologists had difficulty interpreting the CLAHE presentations. These difficulties were evident during the trials as well and are discussed there.

4.2 Trials
The three major questions addressed were (1) are CLAHE and IW equivalent, (2) is CLAHE useful as an adjunct, and (3) are CLAHE readings faster? In addition to the statistical data analysis of the last section, significant information was obtained from the radiologist's comments. The above three questions are viewed in light of the radiologist's comments on the readings forms and their oral discussions with the experimenter. Additionally the radiologist's answers to the exit interview questions are listed.

CLAHE and IW equivalent?
The statistical data analysis indicated that while there was not significant disagreement between CLAHE and IW when examining larger groups of findings, there were differences indicated for some of the smaller groups of findings, specifically those where the most disease occurred. These results are borne out by the radiologists comments. Both verbally and on the score sheets the radiologists indicated they had difficulty with certain clinical findings or comprehension tasks when reading CLAHE. Additionally, when the data was compressed into the 2x2 grid from the 5x5 grid the largest effect was the combining of slightly different CLAHE scores; for instance, combining abnormal and probably abnormal into the same score in the 2x2 grid. This distinction was also evident on the scoresheets where the radiologists recorded seeing the finding with CLAHE, but indicated that they did not feel as comfortable reporting it as they did with IW.

Observation and review of the score sheets indicated that several specific areas were problematic for the observers when reading CLAHE presentations. The areas posing the most difficulty are listed below.

1) Inability to distinguish between soft tissue and fat in mediastinum area because the relationship between original intensity values is lost. This is due to CLAHE modifying the intensity values adaptively on a local scale. Because of this, comparisons that depend on the same relationship between intensity values will not be valid after processing with CLAHE. This is inherently a problem due to the CLAHE technique.
2) Inability to distinguish between vessels and nodes in mediastinum area. Vessels that move through the 3D space axially and are captured on 2D slices are mentally followed through 2D slices by matching intensities to determine where a specific vessel is on the next slice. After CLAHE processing it is more difficult to track the vessel through the slices or to distinguish it from a nodule that may appear on one or

a few slices. While the radiologists had difficulty with this, more experience with CLAHE and following the structure more as opposed to the intensity values may overcome these difficulties. Also, finer slicing on newer CT machines may improve tracking by reducing ambiguities.

3) The lung walls appear thicker. One artifact of CLAHE is that it may slightly thicken edges in an image. Radiologists were unable to diagnosis certain conditions where there appeared to be thickening of the lung wall. Although this may be due in part to the CLAHE processing it is also caused by the radiologists being trained on lung IW images that clip the original information resulting in the low contrast information next to the lung wall not being visible. It is not clear whether making this low contrast area visible with CLAHE will be useful.

4) Lung vessels and nodules stand out more and appear larger under CLAHE. Most of the radiologists adapted to this during the study.

5) The increased markings in the lung due to the enhancement of vessels not well seen under IW caused the radiologists some difficulty in adjusting to the overall busyness of the lung area. In at least one case a radiologist missed some nodules on the CLAHE presentation by simply overlooking them. The radiologist did not see them on the initial readings with CLAHE, saw them on the second reading with IW added, and then saw them clearly on CLAHE when revisited. Most of the radiologists adapted to this during the study.

6) Injected contrast is more difficult to pick out. Contrast is much brighter in IW and stands out better. After CLAHE many areas have been brightened so the injected contrast is less distinguishable. This is inherently a problem due to the CLAHE technique.

7) Landmarks are more difficult to pick out. Most existing landmarks are chosen by their standing out from the background. With CLAHE everything stands out, and intensity based landmarks may be difficult to find. Retraining to use landmarks that stand out structurally rather than through a difference in intensity could solve this problem.

CLAHE as an adjunct?

The statistical data analysis indicates that having CLAHE in addition to IW did not significantly change the radiologists clinical findings. Analysis of the subgroup and individual findings did not point out any statistically significant areas of difference either. Thus having the additional presentation of CLAHE in addition to normal IW views did not change the clinical findings reported. Comments from the radiologists indicated several areas of possible advantage, and some associated disadvantages.

All of the radiologists felt that there was more visualization of structure in the lung. Two of the radiologists commented that it reminded them of high resolution CT (where 1-5mm slice intervals are used instead of 10mm intervals). The disadvantage associated with this was that it took longer to read the area because of the increased markings. Additionally, the radiologists were concerned about overlooking findings in the lung area. While their concerns were real, these concerns are shared with high resolution CT or anything that provides more information to be assimilated by the radiologist. If the additional information is worth the effort, the radiologist adapts to utilize it.

The radiologists also were impressed with the visualization of the liver. A liver window was not provided in the IW views, and in several cases the radiologists felt the liver could be seen better on the CLAHE window. In one case the radiologist saw detail in the liver that he could not see on a liver IW window (provided separately during a training run).

Although the radiologists felt that the above areas merited more investigation to establish whether or not CLAHE could improve clinical readings, they were in agreement that the only reason they would consider using CLAHE views would be as supplements where they might provide better clinical information than was obtainable on the IW views.

CLAHE readings faster?

In the data analysis we found that CLAHE reading times were faster than IW readings times. Observations of the radiologists readings habits strongly supports these conclusions. Significant time is spent

manipulating the films themselves (placing them on the lightbox, repositioning them, moving one's head to view different images). The CLAHE presentation reduces the amount of work required for all these tasks by a factor of two (three if one considers liver windows as well). Additionally, there is less eye and head movement because the radiologist does not need to move back and forth to compare two different IWs of the same anatomical location. This would be especially important for electronic workstations with limited screen space and delays associated with scrolling movements through the image set.

Because CLAHE was not shown to be equivalent to IW for at least some groups of clinical findings, the result that CLAHE readings are faster is not meaningful. For instance, it may be that the CLAHE readings were faster because there was less information in them. Additionally, in the exit interview we find that none of clinicians felt CLAHE presentations alone were adequate. Thus, until we can establish that CLAHE readings are equivalent to IW and acceptable to the clinicians, there is limited significance to the fact that CLAHE reading times are quicker.

Interview
At the conclusion of the study each of the radiologists were interviewed to obtain additional feedback. As part of the interview they were asked to answer five specific questions with regard to reading chest CT studies. The five questions and the radiologists answers are as follows:

(1) *Would CLAHE by itself be a satisfactory clinical presentation (instead of IW) for chest CT?* All the radiologists responded that CLAHE alone would not be a satisfactory clinical presentation because of the problems listed in 4.2.1.
(2) *Would CLAHE be useful as the primary presentation on an electronic workstation where at the push of a button you could switch between CLAHE, and any IW presentation?* One radiologists felt the it might be useful (see advantages below) as long as IW views were available at the push of a button. The other two felt that it would not be useful. Additionally, one radiologist commented that having the initial presentation being CLAHE would not be an advantage because he felt he would need to view all of the regular IW presentations anyway, so having the initial presentation be CLAHE would just consume more time.
(3) *Would CLAHE be a useful adjunct to IW? For instance, on an electronic workstation where it was available at the push of at button?* One radiologist felt it might be useful because of the advantages listed below in (4A). The other two felt it would only be useful if it provided better clinical information than was obtainable on the IW views.
(4) *What are the significant advantages and disadvantages of using CLAHE compared to IW?*

Areas where radiologists responded that CLAHE may have advantages
A) More efficient display technique versus multiple IW presentations. This may be especially advantageous for electronic workstations as economy of screen space and the need to reduce scrolling operations are important issues. Efficiency was due to:
    -less images to view
    -less manipulation of films
    -faster load time
    -less eye and head movement required
    -seeing everything at once, yielding closer to one pass through images

B) Better ability to diagnosis situations requiring viewing the same anatomy with multiple IWs. This is difficult with IW presentations, while with CLAHE the information of all IW settings is viewable on same image at once. This is especially useful when a desired IW setting is not available. In several cases radiologists saw more detail in the bone and liver on CLAHE presentations than they could with IW (lung and mediastinum) presentations.

C) Improved lung detail, similar to high resolution CT images.

D) Improved liver detail compared to soft-tissue window, possibly better than liver IW in some cases.

Areas where radiologists responded that CLAHE may have disadvantages
Replies listed the same disadvantages described in 4.2.1.

(5) *Name the areas where you had difficulty using CLAHE. Given sufficient training (for instance a residency in CLAHE presentations), do you think any of these areas would still be a problem?* Radiologists listed the areas already discussed under disadvantages. All radiologists felt that if the problem was due to CLAHE modifying the relationships between the original intensity values, then more training would not help. For the other types of problems they felt more experience with CLAHE and better correlation to anatomy via pathology, etc. would help.

5. CONCLUSIONS

Three major questions were addressed. First, could we consider CLAHE as equivalent to IW and thus a replacement in the clinic? Second, would CLAHE be useful as an adjunct to IW in the clinic? Finally, is reading with CLAHE faster than with IW?

We found that overall there was not a statistically significant difference between CLAHE/IW and IW/CLAHE, implying that they may be equivalent. However, realizing that most of the agreement came from agreement on normal cases that constituted the majority of the cases, we investigated subgroups where there was more disease. Here we found that for the areas where we had the most disease, we also had more disagreement between CLAHE/IW than IW/CLAHE. CLAHE followed by IW would often demonstrate that the radiologists changed how they scored a clinical finding with CLAHE once they saw the IW presentation. Additionally, when the readers were separately analyzed, one of them showed better agreement for IW/CLAHE than for CLAHE/IW. This implied that like the results from disease subgroups, more changes were made to the CLAHE readings when IW followed, than for when IW was first and CLAHE followed. This statistical evidence correlates with the unanimous agreement among the radiologists that CLAHE alone would not be an acceptable replacement for IW in the clinic. The radiologists exhibited common difficulties in reading CLAHE presentations. Most of the difficulties were related to CLAHE's modifying the relationships between the original intensity values of the images. Most of these difficulties, but not all, could possibly be overcome by adjustment of the radiologists' technique (for instance to use structure instead of intensity cues) and by further training.

There was not a statistically significant difference between the clinical findings of IW only readings and IW with CLAHE as an adjunct. Thus no improvement in adding CLAHE presentations to IW presentations was shown. The radiologists felt that if some of the areas where CLAHE appeared to provide better information could be established, then the CLAHE views would be useful as an adjunct. One radiologist suggested that CLAHE's ability to present all the information in a single view was helpful and may provide benefit as the initial view, as long as IW views were available as well. This was specifically noted for case of electronic workstations where limited screen space makes the advantage of a single presentation more important.

CLAHE was expected to be faster for readings because only one view of the images is necessary while IW presentations require a view for each intensity window and level setting. The reading times for CLAHE were found to be statistically significantly less than IW reading times. Unluckily, for this result to be useful we would need to know that CLAHE and IW presentations were equivalent and that the clinicians would find CLAHE acceptable. Since neither of these were true, there was limited significance to the fact that CLAHE reading times were faster.

6. FUTURE WORK

One of the major problems with using CLAHE alone as the presentation method is that it modifies the relationship between the original intensity values. To overcome this will require investigation of new and different techniques for answering clinical questions that do not depend on the absolute intensity value relationships.

The second major source of problems was lack of experience with CLAHE and the correlation between it and anatomical truth. In this area we are now investigating the specific case of CLAHE presentations of lung parenchyma and the lung wall area by correlating pathology with IW and CLAHE presentations. We hope to find that with enough training and experience the radiologists can become as comfortable reading CLAHE presentations as they are reading IW presentations of these cases.

Finally, we are currently completing development of a high resolution rapid access electronic workstation design. We plan to further investigate whether CLAHE may be useful in combination with IW presentations when each presentation is available at the push of a button.

7. ACKNOWLGEMENTS

8. REFERENCES

1. Pizer SM, Zimmerman JB, Staab EV, "Adaptive Grey Level Assignment in CT Scan Display", *Journal of Computer Assisted Tomography*, 8(2) 300-305, 1984.
2. Pizer SM, Johnston RE, Eriksen JP, "Contrast-Limited Adaptive Histogram Equalization: Speed and Effectiveness", *Proceedings of First Conference on Visualization in Biomedical Computing*, May, 1990.
3. Cromarti R, Pizer SM, "Edge-Affected Context for Adaptive Contrast Enhancement", *Proceedings of the 12th International Conference on Information Processing in Medical Imaging* (IPMI), 474-485, 1991.
4. Zimmerman JB, Pizer SM, Staab EV, Perry JR, McCartney W, Brenton B, "An Evaluation of the Effectiveness of Adaptive Histogram Equalization for Contrast Enhancement", *IEEE Transactions on Medical Imaging*, 7(4), 304-312, 1988.
5. Rehm K, Seeley GW, Dallas WJ, Ovitt RW, Seeger JF, "Design and Testing of Artifact-Suppressed Adaptive Histogram Equalization: A Contrast-Enhancement Technique for Display of Digital Chest Radiographs", *Journal of Thoracic Imaging*, 5(1) 85-91, 1990.
6. Cromarti R, Pizer SM, Rosenman J, Roe CA, Muller K, "Edge-Limited contrast enhancement and its clinical effectiveness for radiation therapy portal films", UNC TR#91-000, submitted to IEEE Transactions on Medical Imaging, 1992.
7. Johnston RE, Yankaskas BC, Perry JR, Pizer SM, Delany DJ, Parker LA, "Agreement Experiments: A Method for Quantitatively Testing New Medical Image Display Approaches", *SPIE Proceedings of Medical Imaging*, Vol 1234, Part II, 1990.
8. Beard DV, "Designing a Radiology Workstation: A Focus on Navigation During the Interpretation Task", *Journal of Digital Imaging*, Vol. 3, No 3, pp 152-163, August 1990.
9. Beard DV, Perry JR, Muller K, Misra R, Brown P, Hemminger BM, Johnston RE, Mauro M, Jaques P, Schiebler M, "Evaluation of Total Workstation CT Interpretation Quality: A Single-Screen Pilot Study", *SPIE Proceedings of Medical Imaging*, Vol 1446, pp 52-58, Feb 1991.
10. Landis JR, Kock GG, "The measurement of observer agreement for categorical data", *Biometrics*, 33, pp 159-174, 1977.