THE CLINICAL PERFORMANCE EVALUATION OF TABLET, LARGE SCREEN TV AND MEDICAL GRAD MONITORS FOR TELERADIOLOGY AND GENERAL USE PURPOSES

by

Roya Nouri Rikabad

B.Sc in Nuclear Physic, Oroumieh University, 2008

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University
2015

THE CLINICAL PERFORMANCE EVALUATION OF TABLET, LARGE SCREEN TV AND MEDICAL GRAD MONITORS FOR TELERADIOLOGY AND GENERAL USE PURPOSES

APPROVED BY:

Assoc. Prof. Albert Güveniş .....
(Thesis Advisor)

Assoc. Prof. Mehmet Burçin Ünlü .....

Assist. Prof. Bora Garipcan .....

DATE OF APPROVAL: 2 JUNE 2015

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere appreciation to my thesis advisor Assoc.Prof.Dr. Albert Güveniş for learning me to how to research as well as for this continuous guidance, support and encouragement that motivated me for the best.

I would like to thank the Radiology Department of GOZTEPE Hospital for their support and contribution through the evaluation period of the study. I specifically appreciate the efforts of Attending Radiologist Dr. Murat Aşık for her continuous support and kind approach as the coordinator of the project and also to Attending Radiologist Dr. Fikret Balyemez and Assistant attending radiologist Dr. Ercan Ayaz.I am fully grateful to them who evaluated the date set in three types of display devices in their tight and stressful schedule and contributions and involvement helped me reach to the final phase of the thesis study.

My thanks and gratitude deep from my heart goes to my parents and my friends Altay Brusan, Hadi Kafil, Sabra Rostami, Meysam Alizad and Seher Uğurcuklu without their support and motivation, it has become so difficult

ABSTRACT

THE CLINICAL PERFORMANCE EVALUATION OF TABLET, LARGE SCREEN TV AND MEDICAL GRAD MONITORS FOR TELERADIOLOGY AND GENERAL USE PURPOSES

One of the most important points in the whole process in digital radiology is to have high quality displays due to deteriorate any information received data acquisition and image processing phases. Nowadays the common display device in observing medical imaging is medical grade TFT LCD monitor. In diagnostic radiology, medical monitors are mainly recommended because of their higher luminance and better contrast ratio. However; the remarkable problem with medical grade monitors is its high cost. In order to find the solution for decreasing the cost; in this study we want to evaluate the clinician's performance of the other display devices such as large screen TV and tablet for observing the medical imaging. Another objective is to evaluate the potential of tablet as Teratological tool for assessing chest X-ray images with nodule.

In this experiment, the data set consisting of 60 chest radiographs were assessed by three experienced radiologists. The area under curve (AUC) of each ROC curve was used as a metric for detecting lung modules in the radiographs. AUC for medical monitor for viewer 1, 2 and 3 were calculated as 0.634, 0.703 and 0.755 respectively. AUC for Tablet for observer 1, 2 and 3 were calculated as 0.634, 0.703 and 0.755 respectively. AUC for large screen TV for radiologist 1, 2 and 3 were calculated as 0.634, 0.703 and 0.755 respectively. According to Analysis of variance or ANOVA test with 95 % confidence interval, there is statistically not significant differences between Medical monitor, tablet, and Large screen TV. Consequently, it is possible to implement Tablet and large screen TV as a medical monitor for medical diagnosis purposes without sacrificing any diagnostic value.

Keywords: Digital Radiology,ROC curve, ANOVA, AUC, Teleradiology

ÖZET

TABLET,BÜYÜK EKRAN TV VE TIBBI GRAD MONITÖRIN TELERADIOLOGY VE GENEL KULLANIM AMACLI KLINIK PERFORMANS DEĞERLENDIRMESI

Dijital radyoloji tüm süreçi içinde en önemli noktalardan biri herhangi bir bilgi kaybı olmadan yüksek kalitede görüntüler sağlamaktır ve bunun için çok yüksek kalitede görüntüleme cihazının olması gerekiyor. Günümüzde tıbbi görüntüleme için kullanılan görüntüleme cihazı tıbbi sınıf TFT LCD monitor dur. Radyolojide, tıbbi monitörler yüksek parlaklık ve daha iyi kontrast oranına sahip oldukları için tavsiye edilir. Tibbi monitörler yüksek maliyetidir ve bunun için bir çozum bulmak amacıyla bu çalışmada tıbbi görüntüleme gözlemlemek için büyük ekran televizyon ve tablet gibi diğer görüntüleme cihazlarının performansını değerlendireceğiz. Diğer amacımız nodül göğüs röntgen görüntüleri değerlendirmek için tabletın potansiyelini Teleradiological bir cihaz olarak tablet ölçmektir.

Bu deneyde, 60 göğüs röntgeni oluşan veri seti, üç deneyimli radyologlar tarafından değerlendirildi. Her bir ROC eğrisinin (AUC) altındaki alan radyografileri akciğer modülleri tespit etmek için bir ölçü olarak kullanıldı. Izleyici 1, 2 ve 3 için tıbbi monitör için AUC sırasıyla 0.634, 0,703 ve 0,755 olarak hesaplandı. Gözlemci, 1, 2 ve 3 için Tablet AUC, sırasıyla 0,634, 0,703 ve 0,755 olarak hesaplanmıştır. Radyolog 1, 2 ve 3 için büyük ekran TV için AUC sırasıyla 0.634, 0,703 ve 0,755 olarak hesaplandı. 95 güven aralığı uymayacak veya ANOVA testi Analizi göre istatistiksel Tıbbi monitör, tablet ve Büyük ekran TV arasında önemli farklar yoktur.

Sonuç olarak, herhangi bir tanısal değerinin ödün vermeden tıbbi tanı amaçlı tıbbi monitör olarak Tablet ve büyük ekran TV uygulamak mümkündür.

Anahtar Sözcülükler: Dijital Radiology,ROC Eğrisi, ANOVA, AUC, Teleradiology

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii
ABSTRACT iv
ÖZET v
LIST OF FIGURES vii
LIST OF TABLES viii
LIST OF ABBREVIATIONS x
1. INTRODUCTION 1
1.1 Background and Motivation 1
1.2 Objective 3
1.3 Outline of the Thesis 3
2. ROC, ANOVA AND KAPPA CHARACTERISTIC AND THEIR APPLICATIONS 4
2.1 The brief review of Receiver Operating Characteristic (ROC) curve . . 4
2.2 Analysis of variance (ANOVA) 11
2.3 Kappa 12
3. ASSESSMENT OF DISPLAY PERFORMANCE FOR MEDICAL IMAGING SYSTEMS 13
3.1 Introduction 13
3.2 Test Patterns 14
4. DESIGN OF EXPERIMENT AND CLINICAL TEST RESULTS WITH STATISTICAL EVALUATION 17
5. DISCUSSION AND CONCRETE RESULTS 47
REFERENCES 51

LIST OF FIGURES

Figure 2.1	ROC Curve	5
Figure 2.2	The distribution of the analysis outcomes	5
Figure 2.3	Criterion value	7
Figure 2.4	The overall diagnostic performance of the test: Four ROC curves with different values of the area under the ROC curve	8
Figure 2.5	A plot of test sensitivity (y coordinate) curve that is versus its false positive rate (x coordinate).	9
Figure 2.6	The empirical ROC curve. The discrete point on the empirical ROC curve are marked with dots.	10
Figure 2.7	The fitted or smooth ROC obtained by the estimation with the at each cutoff level assumption of a binormal distribution	10
Figure 3.1	The TG18-QC pattern	15
Figure 4.1	ROC Curve for Radiologist 1 with Medical Monitor	38
Figure 4.2	ROC Curve for Radiologist 1 with Tablet	39
Figure 4.3	ROC Curve for Radiologist 1 with Large Screen TV	39
Figure 4.4	ROC Curve for Radiologist 2 with Medical Monitor	40
Figure 4.5	ROC Curve for Radiologist 2 with Tablet	41
Figure 4.6	ROC Curve for Radiologist 2 with Large Screen TV	41
Figure 4.7	ROC Curve for Radiologist 3 with Medical Monitor	42
Figure 4.8	ROC Curve for Radiologist 3 with Tablet	43
Figure 4.9	ROC Curve for Radiologist 3 with Large Screen TV	43

LIST OF TABLES

Table 2.1	The Decision Matrix, Sensitivity and Specificity	6
Table 4.1	Features of Medical Monitor, Tablet and Large screen TV	18
Table 4.2	The data set with each radiograph's nodule types	21
Table 4.3	Nodule Coordinates in each radiograph	23
Table 4.4	Radiologist 1's evaluation on Medical Monitor, Tablet, and Large screen TV	27
Table 4.5	Results from Radiologist 1's evaluation on Medical Monitor; Tablet and Large Screen TV	29
Table 4.6	Radiologist 2's evaluation on Medical Monitor, Tablet, and Large screen TV	30
Table 4.7	Results from Radiologist 2's evaluation on Medical Monitor; Tablet and Large Screen TV	33
Table 4.8	Radiologist 3's evaluation on Medical Monitor, Tablet, and Large screen TV	34
Table 4.9	Results from Radiologist 3's evaluation on Medical Monitor; Tablet and Large Screen TV	37
Table 4.10	Statistical Comparison of Radiologist 1's evaluations on Medical Monitors, Tablet and Large Screen TV	38
Table 4.11	Statistical Comparison of Radiologist 2's evaluations on Medical Monitors, Tablet and Large Screen TV	40
Table 4.12	Statistical Comparison of Radiologist 3's evaluations on Medical Monitors, Tablet and Large Screen TV	42
Table 4.13	Diagnostic performance (Area under the ROC Curve) of three observers for the diagnosis of nodule in chest X-ray radiographs on Medical Monitor, Tablet, and Large Screen TV through DBM-MRMC software	44
Table 4.14	95 confidence intervals and hypothesis tests (H0: difference = 0) for treatment AUC differences	44

Table 4.15	Intra-observer agreement for diagnosis of nodule on chest x-ray on the Tablet and Medical monitor	45
Table 4.16	Intra-observer agreement for diagnosis of nodule on chest x-ray on the Large Screen TV and Medical monitor	45
Table 4.17	Total time-consuming and mean time	46
Table 5.1	Radiologist 1 point of view	49
Table 5.2	Radiologist 2 point of view	49
Table 5.3	Radiologist 3 point of view	50

LIST OF ABBREVIATIONS

TFT	Thin Film Transistor
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
CI	Confidence Interval
TPF / TPR	True Positive Fraction / Ratio
FPF / FPR	False Positive Fraction / Ratio
PACS	Picture Archiving and Communication System
AMLCD	Active Matrix Liquid Crystal Display
ANOVA	Analyze of Variance
DICOM	Digital Imaging and Communications in Medicine
NEMA	National Electrical Manufacturers Association
LCD	Liquid Crystal Display

1. INTRODUCTION

1.1 Background and Motivation

Digital radiology can make known as the remarkable technological advancement in medical imaging over the last decades. In digital radiology, Images can be instantly captured, eliminated, modified, and finally delivered to a network of computers. Progressively, medical imaging and patient information are being handled using digital data during receiving, communication, storage, display, analysis, and examination [1].

Picture Archiving and Communication System (PACS) is required in order to manage these large volumes of digital data. PACS is a system for the storage and regulation of medical images. It acts as the file room, reading room, duplicator, and courier. It consists of a networked group of computers; servers; and archives to manage and to store digital images [2]. It can receive any image that is in The Digital Imaging and Communications in Medicine (DICOM) standard format. DICOM was generated by the National Electrical Manufacturers Association (NEMA) to help the distribution and viewing of medical images. A single DICOM file consists of both a header (which stores information about the patient's name, the type of scan, image dimensions, etc), as well as all of the image data [3](which can contain information in three dimensions).

One of the most important points in the whole process in digital radiology is to have high quality displays due to deteriorate any information received data acquisition and image processing phases. The technical, financial, and practical benefits of digital radiology can be entirely exploited only when image analysis can be perfectly done. Mainly, this depends on the performance and quality of display device.

CRT displays were once the only devices which could be used for displaying digital radiographic images. This position changed in the early 21st century following the advancement of Liquid Crystal Displays (LCDs) for use in digital television as well

as in home and business computers [4]. The design and structure of medical monitors have benefitted from these advancements. However it should be considered that digital radiography displays need excellent performance because of the approximately large number of pixels in CR(Computed radiography) and DR (Digital Radiography) images. An LCD is a two-dimensional, electro-optical light modulator that is mounted in front of a back-light. The light is modulated for each pixel by employing an electric field to a thin layer of Nematic liquid crystal mounted between two Polarising films. Active Matrix Liquid Crystal Displays (AMLCDs) make use of the electric field to each pixel utilizing a large array of Thin Film Transistor (TFT) switches made from amorphous silicon (a-Si) deposited on a glass substrate. On the other hand, the development of such large TFT arrays has also led to the subsequent development of DR image receptors [5].

Consequently, nowadays the common display device in observing medical imaging is medical grade TFT LCD monitor. In diagnostic radiology, medical monitors are mainly recommended because of their higher luminance and better contrast ratio [6].

Since TFT LCD Panel technology advanced very quickly in last decade from performance and technical improvement point of view, the quality level of images obtained by using medical grade TFT LCD monitors could be currently comparable to the quality level of images acquired by using consumer grade TFT LCD monitors. Due to some articles which pay attention to the comparison of medical grade LCD monitors commonly with consumer grade LCD monitors, the consumer grade monitors with some special technical qualification could be also applied as a PACS monitor for medical diagnosis purpose [7].

The noticeable problem with medical grade TFT LCD monitors is its high cost. The cost of a medical monitor can be at least ten times higher than the cost of regular grade monitors. On the other hand; it is difficult to use medical monitors in the many contexts of Teleradiology. For example the radiologist may be at home or traveling [6].

In order to find the solution for decreasing the cost; in this study we want to

evaluate the clinician's performance of the other display devices such as large screen TV and tablet for observing the medical imaging. We want to examine the potential of the Tablet and even Large Screen TV as a teleradiologic tool for evaluating chest x-ray images with nodules [8]. The comparison of radiological's performance was undertaken through detecting nodule in chest x-ray images by large screen TV and Tablet. The comparison between MEDICAL GRADE TFT LCD MONITOR; LARGE SCREEN TV and TABLETs was targeted to be realized through receiver operating characteristic (ROC) curve analysis in a clinical experiment.The ROC curve is a fundamental tool for diagnostic test evaluation. The diagnostic performance of a test, or the accuracy of a test to differentiate diseased cases from normal cases is assessed using Receiver Operating Characteristic (ROC) curve analysis. ROC curves can also be used to the diagnostic performance of two or more laboratory or diagnostic tests [9].

1.2 Objective

The first aim is to evaluate the Tablet and Large Screen TV performance to be as compatible as possible to a Medical Grade Monitor performance with noticeable cost benefit. The second aim is to examine their potential as the Teleradiologic tool. The validity of the experiment was achieved to be realized through receiver operating characteristic (ROC) curve analysis by a clinical evaluation method.

1.3 Outline of the Thesis

Chapter 1 proposes the subject and determines the structure of the study. The receiver operating characteristic curves and the applications are explained in chapter 2. The quality control and calibration method for display systems are presented in chapter 3. The method of experiment and clinical test results is placed in Chapter 4. The future work and discussion about results are presented in Chapter 5.

2. ROC, ANOVA AND KAPPA CHARACTERISTIC AND THEIR APPLICATIONS

2.1 The brief review of Receiver Operating Characteristic (ROC) curve

The ROC curve was first applied during World War II for the analysis of radar signals before it was used in signal detection theory. In 1941, radar receiver operators were being assessed on their ability to differentiate signal from noise. In the 1950s, ROC curves were employed in psychophysics to assess human detection of weak signals. Since that time, ROC analysis has been used in a number of fields including engineering, quality control, and weather forecasting. Its use in medicine to assess diagnostic test performance was first described by Lusted in 1971 [10].ROC curves can be used to compare the diagnostic performance of two or more laboratory or diagnostic tests. It is become obvious that they are remarkably useful in medical decision-making. Receiver Operating Characteristic (ROC) curve analysis is used to evaluate accuracy of a test to discriminate diseased cases from normal cases. With other definition, it can be used to evaluate the diagnostic performance of a test and to determine a cutoff value for a clinical test [11].

The ROC curve can be represented equivalently by plotting the fraction of true positives (TP = true positive rate) vs. the fraction of false positives (FP = false positive rate).ROC graphs are two-dimensional graphs in which TP rate is plotted on the Y axis and FP rate is plotted on the X axis. With the other point of view, an ROC graph describe relative tradeoffs between benefits (true positives) and costs (false positives) and represent a graph of sensitivity (y-axis) vs. 1 – specificity (x-axis) [12].In other words, in a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points of a parameter as shown in Figure 2.1.Each point on the ROC curve determines a sensitivity/specificity pair matching to a specific decision threshold. [13]

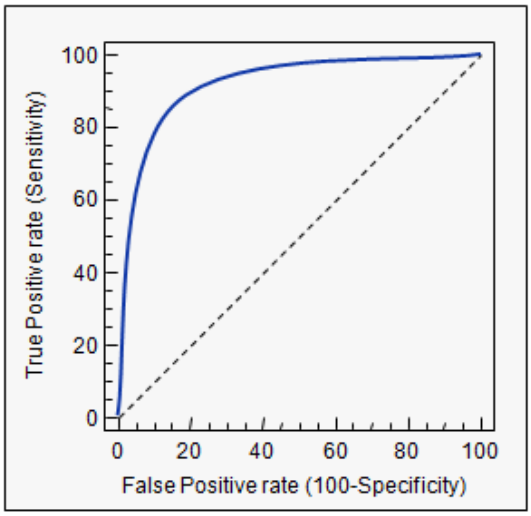


Figure 2.1 ROC Curve

When it is considered the outcomes of a particular test in two groups, one group with a disease, the other group without the disease, it will be barely recognized a perfect separation between the two groups [14]. Certainly, the distribution of the analysis outcomes will overlap, as shown in Figure 2.2.

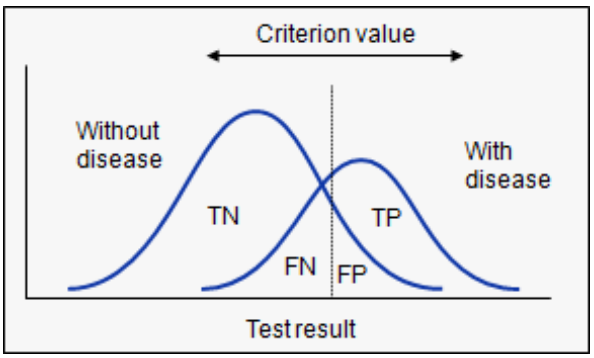


Figure 2.2 The distribution of the analysis outcomes

For each possible cut-off point or criterion value it is preferred to discriminate between the two groups, some cases with the disease will be correctly classified as positive (TP = True Positive fraction), but some cases with the disease will be classified negative (FN = False Negative fraction). On the other point of view, There will be some cases without the disease accurately classified as negative (TN = True Negative fraction), but some cases without the disease will be classified as positive (FP = False Positive fraction) as shown in Figure 2.2 And Table 2.1 [15].

Table 2.1
The Decision Matrix, Sensitivity and Specificity

DISEASE					
TEST	PRESENT	N	ABSENT	n	TOTAL
POSITIVE	True Positive(TP)	A	False Positive	c	a+c
NEGATIVE	False Negative(FN)	B	True Negative(TN)	d	b+d
TOTAL		a+b		c+d	

Finally, sensitivity and specificity which comprise the basic measures of performance of diagnostic tests can be defined according to Table 2.1 .Sensitivity is the probability that a test result will be positive when the disease is present (true positive rate, expressed as a percentage) that is equal with (a / (a+b)). Specificity is the probability that a test result will be negative when the disease is not present (true negative rate, expressed as a percentage) which is equal with (d / (c+d)) [6].

When a higher threshold value is selected, the specificity will increase while the false positive fraction will decrease but on the other hand the true positive fraction and sensitivity will decrease. When a lower criterion value is selected, although the true positive fraction and sensitivity will increase, the false positive fraction will increase. So it is obvious that the true negative fraction and specificity will decrease [10]as shown in 2.2. This determinants that the selection of cut-off point of threshold value is so important in the results of test [6].

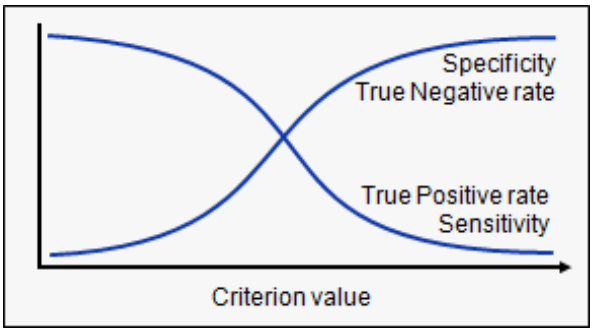


Figure 2.3 Criterion value

Furthermore, one of the most favored measures is the area under the ROC curve denoted as AUC.The area under the ROC curve (AUC) is a measure of how well a parameter can discriminate between two diagnostic groups (abnormal/normal).It is generally used as a global index of diagnostic performance. It can be represented that the AUC is equal to the probability that the viewer will accurately classify the positive case when presented with a randomly chosen pair of cases in which one case is positive and one case is negative.In other words, AUC is a combined measure of sensitivity and specificity [13]. It is generally a measure of the overall performance of a diagnostic test and is illustrated as the average value of sensitivity for all possible values of specificity. Because of the x and y axes values which have values rating from 0 to 1, AUC can take on any value between 0 and 1.When the amount of AUC is closer to 1, the better the overall diagnostic performance of the test is recognized to be correctly accurate as shown by A curve on Figure 2.4.This curve would be clarified as the viewer having a 100 probability of perfectly arranging a random positive–negative case pair. The possible lower limit for the AUC of a diagnostic test is 0.5. The line segment from 0, 0 to 1, 1 has an area of 0,5 as presented by curve D on Figure 2.4 . If the viewer was comprehensively inexperienced and/or the test was entirely unplanned, similar to blind guessing, then the ROC curve would be a straight line connecting the lower left to upper right corners, and the area under this curve would be 0.5. In practice, an obtained curve from diagnostic test result scan be plotted between curve A and curve D such like curve C and B as shown in Figure 2.4 .It represents at least some ability to distinguish between subjects with and without a particular disease. It is obvious that

test B represents better diagnostic performance than test C because of having higher AUC value. [13]

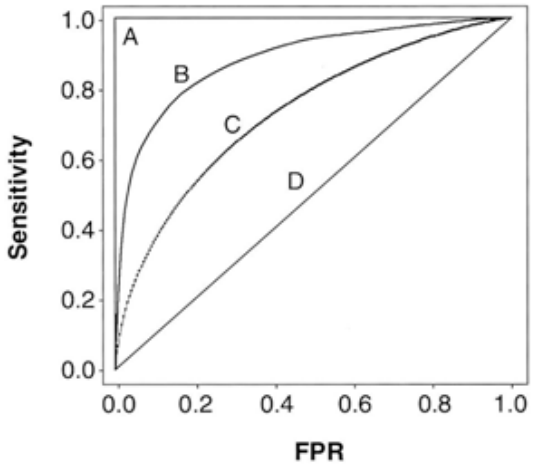


Figure 2.4 The overall diagnostic performance of the test: Four ROC curves with different values of the area under the ROC curve

When the results of a test fall into one of two obviously defined categories, such as either the presence or absence of a disease, then the test has only one pair of sensitivity and specificity values. However, in many diagnostic situations, making a decision in a binary mode is both difficult and impractical.

Finally, a single pair of sensitivity and specificity values is inadequate to characterize the full range of diagnostic performance of a test [10]. For example, if there will be N number of patients with definite problem of pulmonary nodules who got chest radiography to decide whether the problem is present or absent. Chest radiographs could be evaluated according to a five-point scale: 1 (definitely no nodule present), 2 (probably no nodule present), 3 (possibly no nodule present), 4 (probably nodule present), and 5 (definitely nodule present). In this example, one can choose from four different cutoff levels to distinguish a positive test for nodule present on the chest radiographs: X. >2 (the most liberal criteria), >3, >4, and 5 (the most stringent criteria). Therefore, there will be four pairs of sensitivity and specificity values, one pair for each cutoff

level, and the sensitivities and specificities rely on the cutoff levels which are utilized to express the positive and negative test results [16]. As the cutoff level reduces, the sensitivity raises while the specificity goes down, and vice versa. To deal with these multiple pairs of sensitivity and specificity values, one can draw a graph using the sensitivities as the y coordinates and the 1-specificities or FPRs as the x coordinates as shown at Figure 2.5. Each individual point on the graph, called an operating point, is achieved by using different cutoff levels for a positive test result. An ROC curve can be guessed from these discrete points, by making the acceptance that the test results follow a certain distribution. The resulting curve is called the fitted or smooth ROC curve as shown at Figure 2.6. The assessment of the smooth ROC curve relied on a binormal distribution uses a statistical method called maximum likelihood estimation (MLE) [16]. Another way to construct an ROC curve is to attach all the points gained at all the possible cutoff levels and the two endpoints on the ROC curve are 0, 0 and 1, 1 with each pair of values comparable to the FPR and sensitivity, respectively as shown at Figure 2.7. The resulting ROC curve is called the empirical ROC curve. The ROC curve demonstrates the relationship between sensitivity and FPR. Because the ROC curve shows the sensitivities and FPRs at all possible cutoff levels, it can be utilized to evaluate the performance of a test without depending on the decision threshold.

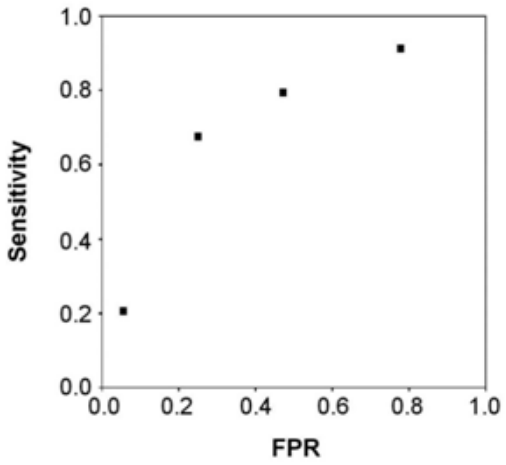


Figure 2.5 A plot of test sensitivity (y coordinate) curve that is versus its false positive rate (x coordinate).

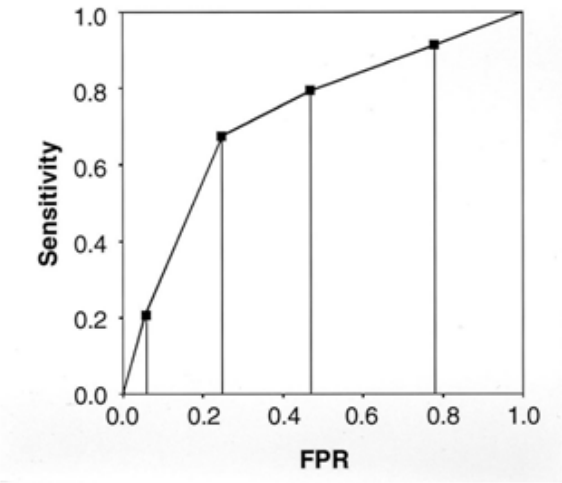


Figure 2.6 The empirical ROC curve. The discrete point on the empirical ROC curve are marked with dots.

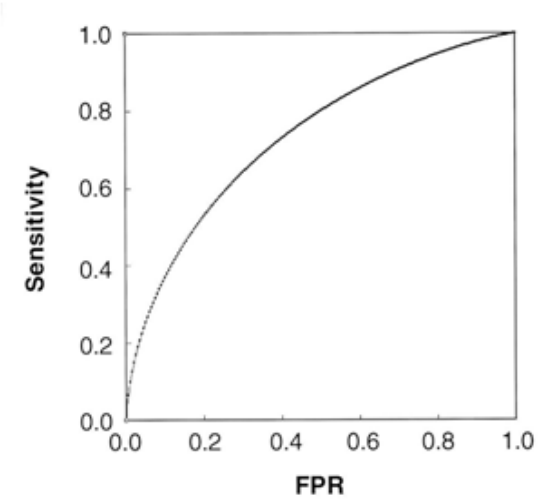


Figure 2.7 The fitted or smooth ROC obtained by the estimation with the at each cutoff level assumption of a binormal distribution

2.2 Analysis of variance (ANOVA)

Analysis of variance (ANOVA) is statistical technique for determining the degree of difference or similarity between multiple independent populations. It is based on the comparison of the average value of a common component between two or more groups.In the simple point of view, ANOVA supplies a statistical test of whether or not the means of several groups are equal and also gives statistical significance which is useful in comparing three or more means (groups or variables) [17].

Analysis of variance is a specific form of statistical hypothesis testing frequently used in the evaluation of experimental data. It analyzes the total variation in a response variable into variability within groups and variability between groups. It is utilized to estimate group means and standard errors and to evaluate magnitude of variation attributable to specific sources. ANOVA test gives the information about whether there is statistically difference between the evaluated independent groups or not. With the other definition, it is a statistical method that yields values that can be evaluated to determine whether a significant relation exists between variables.There are three type of ANOVA, one-way ANOVA which contains one factor, two-way factors which includes two factors, and finally three-way ANOVA which consists of three factors [18].

A statistical hypothesis test is a method of making decisions using data. A test result which is calculated from the null hypothesis and the sample is called statistically significant if it is deemed unlikely to have occurred by chance, assuming the truth of the null hypothesis. A statistically significant result, when a probability (p-value) is less than a threshold (significance level), justifies the rejection of the null hypothesis, but only if the a priori probability of the null hypothesis is not high. [19]In the typical application of ANOVA, the null hypothesis is that all groups are simply random samples of the same population.

2.3 Kappa

Items such as physical exam findings, radiographic interpretations, or other diagnostic tests often rely on some degree of subjective interpretation by observers. Studies that measure the agreement between two or more observers should include a statistic that takes into account the fact that observers will sometimes agree or disagree simply by chance [20]. The kappa statistic (or kappa coefficient) is the most commonly used statistic for this aim. A kappa of 1 indicates perfect agreement, whereas a kappa of 0 indicates agreement equivalent to chance Inter-observer variation can be measured in any situation in which two or more independent observers are evaluating the same thing [21].The calculation is depended on the difference between how much agreement is actually present ("observed" agreement) compared to how much agreement would be expected to be present by chance alone ("expected" agreement) [20].

3. ASSESSMENT OF DISPLAY PERFORMANCE FOR MEDICAL IMAGING SYSTEMS

3.1 Introduction

The medical image display is usually the last step of a medical imaging succession. Medical images are originally achieved by imaging modalities such as x-ray, ultrasound (US), magnetic resonance imaging (MRI), computed tomography (CT), or nuclear medicine scans .These methods measure physical or functional aspect of the patient in the form of multidimensional data sets. Images differ generally in their attributes such as size, spatial resolution, and data depth. The advent of digital modalities led to the creation of essentially electronic images.

Medical images could be viewed on a video display device with the capability to change the appearance of the image by display workstations. These devices were used generally for achieving and displaying digital images from a few similar imaging instruments, and the image appearance was altered using the "brightness" and "contrast" controls of the display device. The "fluidity" of soft-copy presentation increases the important of the consistency of image appearance. The cross-utilization of both soft-copy and hard-copy images caused to new challenges in consideration to diagnosticians, raising the need for acceptance testing and quality control of electronic medical displays.

In a modern picture archiving and communications system (PACS) environment, images from a number of implement of varying type may be displayed or printed in a difference of locations by different individuals. It is obvious that standards are important to successful assimilation of these components. Standardization must contain the communications protocols, data formats, ability to perform the consistency of image display and presentation among the modalities, printers, and workstations where images will be viewed [22].

According to AAPM standards; there are some parameters which be considered for quality control of display devices in medicine use. The evaluation of Geometric Distortions, Display Reflection, Luminance Response, Luminance Spatial and Angular Dependencies and Display Resolution are the remarkable quality control factors. There are two acceptable ways to assess these factors: Visual Evaluation and Quantitative Evaluation [23].

3.2 Test Patterns

The Visual interpretation of display performance should be operated by definite Test Patterns which are regulated by AAPM. All patterns are supplied in three formats: DICOM, 16-bit TIFF, and 8-bit TIFF. The DICOM and 16-bit TIFF patterns contain 12 bits of pixel values, while the 8-bit TIFF patterns only contain an 8-bit range of pixel values. There are several types of test patterns which construct for diagnosing the specific characterization of display systems such as the TG18-CX;TG18-LPV/LPH;TG18-AD;TG18-CT and TG18-QC. Routine visual evaluations of performance are usually done using a single comprehensive test pattern. A new pattern designed by the AAPM Task Group 18 committee, referred to in this report as the TG18-QC pattern, is confirmed for overall display quality assessment [23]. The TG18-QC test pattern is shown in Figure 3.1.

The explanation of the TG18-QC pattern:

Grid lines (one pixel) with thicker lines (three pixels) along periphery and around central region are used to evaluate of geometric distortions. Geometric distortion can be measured in terms of the quantify of spatial angulation or two dimensional dislocation in a geometric test pattern, and be expressed in terms of pixels, spatial dimensions (i.e., millimeters), or percent differences in various directions or areas [23].

Sixteen 102×102 (1k version) luminance patches with pixel values varying from 8 to 248 (in 8-bit version) [128 to 3968 in 12-bit version] are used to assess luminance

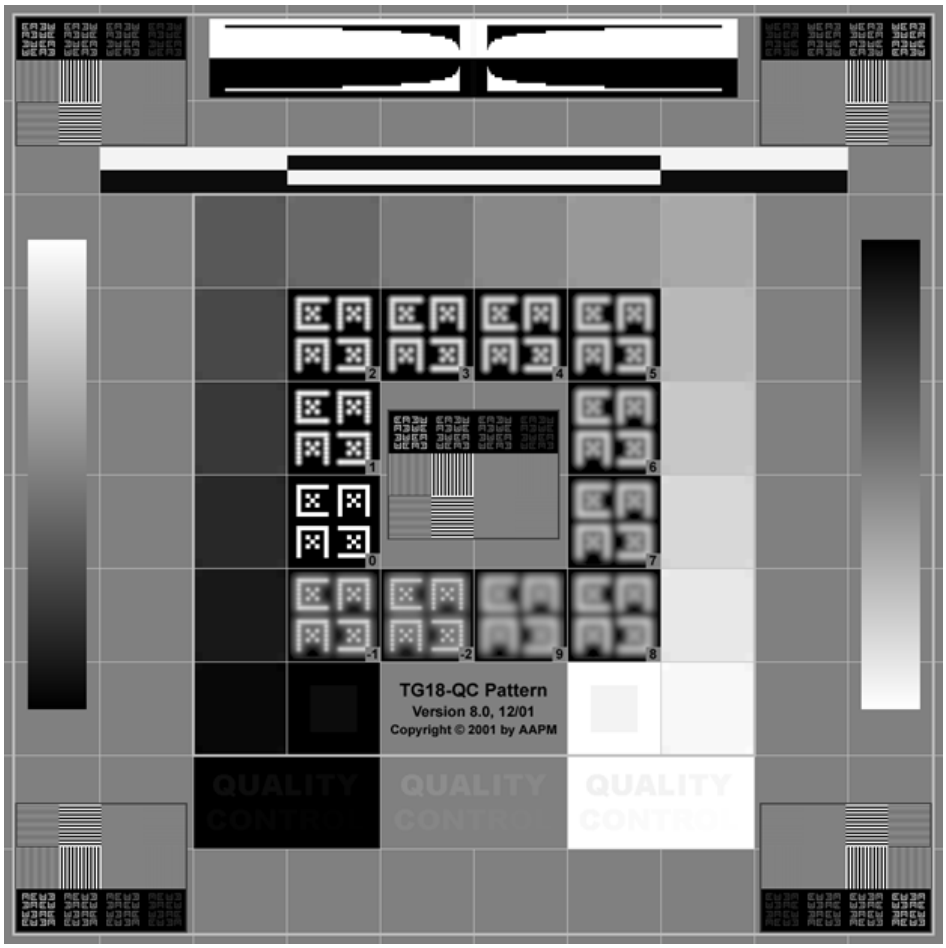


Figure 3.1 The TG18-QC pattern

response. The luminance response of a display device mentions to the relationship between viewed luminance and the input values of a standardized display system. The displayed luminance contains of light produced by the display device that varies between maximum and minimum luminance, along with a fixed contribution from diffusely reflected ambient light [23].

Each patch includes four small 10 × 10 corner patches (1k version) at ±4 [±64] of pixel value difference from the background, +4 [+64] in upper left and lower right, –4 [–64] in lower left and upper right. The small patches are utilized to evaluate visual assessment of luminance response. In addition, two patches with minimum and maximum pixel value are installed containing 13 [205], and 242 [3890] pixel value internal patches, similar to 5 and 95 areas in the SMPTE test pattern [23]. Line-pair patterns at the center and four corners at Nyquist and half-Nyquist frequencies which have having pixel values at 0–255 [0–4095] and 128–130 [2048–2088] are used to assess resolution. The quantitative measure of the ability of a display system to produce separable images of different points of an object with high fidelity is called spatial resolution [23].

In addition, in order to evaluate resolution , “Cx” patterns at the center and four corners with pixel values of 100, 75, 50, and 25 of maximum pixel values against a zero pixel value background are used [23].

For assessing luminance levels Contrast-detail “QUALITY CONTROL” letters with various contrasts at minimum, midpoint, and maximum pixel values are used. On the other hand, for assessing bit depth and contouring artifacts two vertical bars with continuous pixel value variation are used. Furthermore, for evaluating video signal artifacts White and black bars are utilized, similar to those in the SMPTE pattern [23].

4. DESIGN OF EXPERIMENT AND CLINICAL TEST RESULTS WITH STATISTICAL EVALUATION

In this experiment, the data set consisting of 60 chest radiographs was used. Thirty had nodule radiological signs of pulmonary disease, while the other 30 displayed no significant abnormality. This data set has been achieved from the writer of a previous study [16, 24]. These 60 normal digital adult chest radiographs be randomly preferred from a wide range of patients' database who had no pulmonary disease [24]. All patient-related information was said to be digitally obscured from radiographs. Lung nodules with various subject-contrasts and two different diameters (5mm and 10 mm) were mentioned to be reproduced by digitally superimposing circular Gaussian profiles on half of normal radiographs with Matlab software (Mathworks Inc, Natick). In conclusion, a set of 60 radiographs was gained that consisted of 30 normal radiographs and 30 solitary-nodule radiographs having simulated nodules located on different places within the lung fields.

These 60 chest radiographs were independently evaluated by three Radiologists in three types of display devices including Medical Monitor, Tablet, and Large screen TV in order to assess their performance. The technical information about Medical Monitor, Tablet and large screen TV is given in Table 4.1.

Table 4.1
Features of Medical Monitor, Tablet and Large screen TV

Display type	Medical monitor	Tablet	Large Screen TV
Manufacture	NEC	Lenova	Philips
Model	NEC MultiSync® MD213MG	Smart Tab II	40PFK4009/12
Viewable Image Size	21.3”	10” WXGA	40”
Display type	3 MP Grayscale	2 MP TFT Active matrix color LCD	2 MP LED Full HD
Resolution	2048 x 1536 pixels	1280 x 800 pixels	1920x1080 pixels
Contrast ratio	900:1	800:1	1200:1
Maximum Brightness	1450 (cd/m^2)	400 (cd/m^2)	280(cd/m^2)
Minimum Brightness	0.30 (cd/m^2)	0.27 (cd/m^2)	0.21 (cd/m^2)
Used brightness	400 cd/m2 calibrated	220 (cd/m^2)	105(cd/m^2)
Input signal	DVI-D, DVI-I	Dock Connector (Micro USB)	HDMI/USB/DVI/ SCART(RGB/CVBS)
Ambient lighting	27-30 Lux	27-30 Lux	27-30 Lux
DICOM Viewer	Micro DICOM Viewer	Direct DICOM Viewer for Android	Micro DICOM Viewer
Use time	more than 5 years	more than 3 years	more than 3 years

Three radiologists from GÖZTEPE EDUCATION Hospital, GÖZTEPE Istanbul, cooperated in the study. One of them was attending radiologist and the others were 4th-year resident and 3rd-year resident radiologists in Radiology department at GÖZTEPE Education Hospital. They independently assessed the set of images on Medical monitor, Tablet and the Large screen TV. The radiographs were randomly displayed them. They were wanted to rank their level of confidence in the presence or absence of a nodule in the scoring forms by using a rating scale (1- definitely no nodule present, 2-probably no nodule present, 3- possibly nodule present, 4-probably nodule present and 5-nodule definitely present). They were also demanded to submit the location of each suspected nodule to check that the diagnosed nodule coincided to the position of the simulated nodule. During test period, the radiologists were allowed to change in the setting of display systems such as brightness, contrast, zoom, the window level and width. It was told radiologists which viewing time is unrestricted. The time spent for reading each radiograph was recorded separately, and the total time spent for any radiologist in observing all 60 radiographs in each display system was independently recorded. The interruption between sessions was at least 6 weeks to remove the learning effects. The observations were done in the same location and was tried to control ambient light at the same level for all observers in each training session. As Dicom viewer, Micro Dicom Viewer software tool has been used in the evaluation of Medical Monitor and Large screen TV, and for the evaluation of Tablet performance, Direct Dicom Viewer software tool for android systems has been used. The detailed content of the data set with nodule coordinates is represented in Table 4.2 and Table 4.3.

The results of Radiologists' evaluations for 60 radiographs on Medical monitor, Tablet, and Large screen TV are given in Tables 4.4 – 4.9. In order to check quality control of display systems, the visual quality control test was done. TG18-QC TEST pattern which is recommended for overall display quality assessment by AAPM was used [23]. As it is clear, contrast, resolution and gamma are three factors in display systems which should be considered carefully. According to grayscale calibration; grayscale standard display function (GSDF) or gamma 2.2 was set in Medical Monitor. While the Gamma value was set 1.8 in tablet and 2.6 in large screen TV.

JROCFIT that is Johns Hopkins University's specific edition for ROCFIT software package from Chicago University was used to analyze the receiver operating characteristics curve [25]. Through JROCFIT software tool, ROC curves for all radiologists and display devises are showed in Figures 4.1-4.9. It is also used to compare sensitivity and specificity of Medical Monitor, Tablet and Large screen TV. These details are given in Table 4.10- 4.12 .For Analysis of Variance or ANOVA tests , the area under receiver operating characteristic (AUC) for displaying devices were analyzed by the Dorfman-Berbaum-Metz method using RSCORE and semi-parametric estimation of ROC indices with DBM-MRMC software [26]. In determining the differences between overall radiologist's performances, the AUC and P-value were adjusted by the same method and also confidence interval for each one is obtained. Due to ANOVA test, whether there are statistically significant differences between three types of display devices or not is determined. These data is shown in Table 4.13 and 4.14 and figure 4.1-4.9. Furthermore, the intra-observer value or kappa value was calculated through GraphPad software. The results are given in Table 4.15 and 4.16. For each stage of training, time-consuming was recorded. Total time-consuming and mean time are represented in Table 4.17.

Table 4.2
The data set with each radiograph's nodule types

Nodule Type	nodule FWHM (mm)
s = subtle	5 and 10
k = small	5
g = large	10
b = no nodule	

Image Number	Nodule Type	Image Number	Nodule Type
1	k9	16	b29
2	g7	17	b20
3	k13	18	k1
4	g20	19	b6
5	s9	20	k22
6	b21	21	b15
7	g23	22	k28
8	b4	23	b26
9	b13	24	b27
10	s8	25	k16
11	b5	26	k25
12	s4	27	g17
13	b2	28	b12
14	k27	29	s7
15	b18	30	b24

Image Number	Nodule Type	Image Number	Nodule Type
31	b7	46	b17
32	b1	47	b14
33	b8	48	b3
34	b28	49	b25
35	b10	50	g27
36	k20	51	g16
37	b22	52	k12
38	k3	53	b16
39	b11	54	s2
40	b9	55	g21
41	k4	56	k2
42	k15	57	b29
43	k14	58	b23
44	s5	59	s3
45	b19	60	s3

Table 4.3
Nodule Coordinates in each radiograph

COORDINATES			
sign	NODULE(mm)	X-axis	Y-axis
s1	5	404	1032
s2	5	1164	1016
s3	5	1419	542
s4	5	276	1368
s5	5	663	540
s6	10	1111	544
s7	5	1239	1099
s8	5	625	536
s9	5	652	1035
s10	-	-	-
s11	5	352	618

COORDINATES			
sign	NODULE(mm)	X-axis	Y-axis
k1	5	521	470
k2	5	1335	928
k3	5	1128	241
k4	5	252	1115
k5	5	1369	532
k6	5	1312	712
k7	5	675	363
k8	5	1320	880
k9	5	1367	530
k10	5	454	675
k11	5	375	1040

COORDINATES			
sign	NODULE(mm)	X-axis	Y-axis
k12	5	385	1028
k13	5	1302	451
k14	5	502	939
k15	5	1217	1307
k16	5	1213	1142
k17	5	1270	845
k18	5	667	482
k19	5	1438	776
k20	5	1259	490
k21	5	1107	241
k22	5	467	654
k23	5	1238	1277
k24	5	301	1137
k25	5	1329	529
k26	5	506	477
k27	5	1341	481
k28	5	1148	947
k29	5	1459	1150

COORDINATES			
sign	NODULE(mm)	X-axis	Y-axis
g1	10	360	908
g2	10	560	429
g3	10	1283	618
g4	10	1165	422
g5	10	1564	1057
g6	10	1246	554
g7	10	1414	105
g8	10	460	580
g9	10	1385	1094
g10	10	1225	597
g11	10	1200	497
g12	10	1245	411
g13	10	1383	699
g14	10	367	1001
g15	10	522	1278
g16	10	1315	1126
g17	10	1495	1372
g18	10	516	1131
g19	10	1182	345
g20	10	488	599
g21	10	484	554
g22	10	344	1386
g23	10	1285	495
g24	10	563	507
g25	10	459	544
g26	10	477	995
g27	10	1349	632
g28	10	571	465
g29	10	1137	587

Radiograph Number	Type	Nodule Existence in Reality	Rating(1-5) For Medical Monitor	Rating(1-5) For Tablet	Rating(1-5) For Large Screen TV
20	k22	1	4	4	4
21	b15	0	2	2	1
22	k28	1	2	2	2
23	b26	0	1	1	2
24	b27	0	1	2	1
25	k16	1	4	2	2
26	k25	1	3	3	5
27	g17	1	2	3	4
28	b12	0	2	2	2
29	s7	1	1	1	1
30	b24	0	2	2	2
31	b7	0	1	1	1
32	b1	0	2	1	2
33	b8	0	2	5	2
34	b28	0	2	4	2
35	b10	0	3	1	1
36	k20	1	5	5	5
37	b22	0	4	2	1
38	k3	1	3	5	5
39	b11	0	3	4	3
40	b9	0	1	2	4
41	k4	1	4	3	3
42	k15	1	4	1	1

Table 4.4
Radiologist 1's evaluation on Medical Monitor, Tablet, and Large screen TV

Radiograph Number	Type	Nodule Existence in Reality	Rating(1-5) For Medical Monitor	Rating(1-5) For Tablet	Rating(1-5) For Large Screen TV
1	k9	1	4	5	5
2	g7	1	4	4	4
3	k13	1	2	3	2
4	g20	1	5	5	5
5	s9	1	4	1	1
6	b21	0	2	2	2
7	g23	1	2	2	3
8	b4	0	3	3	1
9	b13	0	1	2	2
10	s8	1	5	2	2
11	b5	0	2	4	1
12	s4	1	2	4	1
13	b2	0	1	1	1
14	k27	1	2	2	2
15	b18	0	4	4	2
16	b29	0	3	4	3
17	b20	0	1	1	1
18	k1	1	5	5	5
19	b6	0	3	3	3

Radiograph Number	Type	Nodule Existence in Reality	Rating(1-5) For Medical Monitor	Rating(1-5) For Tablet	Rating(1-5) For Large Screen TV
43	k14	1	2	2	5
44	s5	1	4	4	2
45	b19	0	3	3	3
46	b17	0	1	1	2
47	b14	0	3	2	1
48	b3	0	1	2	2
49	b25	0	3	2	3
50	g27	1	2	4	2
51	g16	1	4	4	4
52	k12	1	5	5	4
53	b16	0	4	2	2
54	s2	1	2	1	2
55	g21	1	1	2	2
56	k2	1	3	4	4
57	b29	0	3	2	3
58	b23	0	3	1	1
59	s3	1	4	5	5
60	s6	1	3	3	3

Table 4.5
Results from Radiologist 1's evaluation on Medical Monitor; Tablet and Large Screen TV

Display Devices	Medical Monitor	Medical Monitor	Tablet	Tablet	Large Screen TV	Large Screen TV
Type of cases	P Cases	N cases	P Cases	N cases	P Cases	N cases
1 (definitely no nodule present)	2	9	4	8	4	11
2 (probably no nodule present)	9	8	7	13	9	12
3 (possibly nodule present)	4	10	5	3	3	6
4 (probably nodule present)	10	3	7	5	6	1
5 (definitely nodule present)	5	0	7	1	8	0
TOTAL	30	30	30	30	30	30

Table 4.6
Radiologist 2's evaluation on Medical Monitor, Tablet, and Large screen TV

Radiograph Number	Type	Nodule Existence in Reality	Rating(1-5) For Medical Monitor	Rating(1-5) For Tablet	Rating(1-5) For Large Screen TV
1	k9	1	4	4	4
2	g7	1	5	3	4
3	k13	1	3	1	2
4	g20	1	5	5	5
5	s9	1	3	1	3
6	b21	0	2	2	1
7	g23	1	3	3	3
8	b4	0	2	1	2
9	b13	0	2	1	1
10	s8	1	4	4	4
11	b5	0	3	1	3
12	s4	1	2	1	2
13	b2	0	3	2	3
14	k27	1	2	1	2
15	b18	0	3	1	1
16	b29	0	3	1	1
17	b20	0	3	1	3
18	k1	1	5	5	4
19	b6	0	4	1	1

Radiograph Number	Type	Nodule Existence in Reality	Rating(1-5) For Medical Monitor	Rating(1-5) For Tablet	Rating(1-5) For Large Screen TV
20	k22	1	4	4	4
21	b15	0	3	1	3
22	k28	1	4	1	1
23	b26	0	1	1	2
24	b27	0	2	1	2
25	k16	1	4	1	4
26	k25	1	2	1	2
27	g17	1	3	3	2
28	b12	0	1	1	1
29	s7	1	3	1	1
30	b24	0	3	1	3
31	b7	0	3	1	1
32	b1	0	2	2	2
33	b8	0	3	2	4
34	b28	0	2	2	2
35	b10	0	3	1	3
36	k20	1	5	5	4
37	b22	0	2	2	2
38	k3	1	4	4	5
39	b11	0	3	4	3
40	b9	0	3	2	1
41	k4	1	2	2	2
42	k15	1	3	3	1

Radiograph Number	Type	Nodule Existence in Reality	Rating(1-5) For Medical Monitor	Rating(1-5) For Tablet	Rating(1-5) For Large Screen TV
43	k14	1	3	3	4
44	s5	1	4	4	2
45	b19	0	3	1	1
46	b17	0	2	1	2
47	b14	0	3	1	1
48	b3	0	1	1	1
49	b25	0	2	3	2
50	g27	1	2	2	2
51	g16	1	4	4	4
52	k12	1	4	4	4
53	b16	0	4	2	1
54	s2	1	4	4	4
55	g21	1	2	2	2
56	k2	1	3	4	4
57	b29	0	4	2	1
58	b23	0	1	1	1
59	s3	1	3	4	3
60	s6	1	3	2	1

Table 4.7
Results from Radiologist 2's evaluation on Medical Monitor; Tablet and Large Screen TV

Display Devices	Medical Monitor	Medical Monitor	Tablet	Tablet	Large Screen TV	Large Screen TV
Type of cases	P Cases	N cases	P Cases	N cases	P Cases	N cases
1 (definitely no nodule present)	0	3	8	19	4	15
2 (probably no nodule present)	6	9	4	9	9	7
3 (possibly nodule present)	10	14	6	1	4	7
4 (probably nodule present)	10	4	9	1	11	1
5 (definitely nodule present)	4	0	3	0	2	0
TOTAL	30	30	30	30	30	30

Table 4.8
Radiologist 3's evaluation on Medical Monitor, Tablet, and Large screen TV

Radiograph Number	Type	Nodule Existence in Reality	Rating(1-5) For Medical Monitor	Rating(1-5) For Tablet	Rating(1-5) For Large Screen TV
1	k9	1	4	5	4
2	g7	1	4	2	3
3	k13	1	3	3	3
4	g20	1	5	5	5
5	s9	1	3	1	1
6	b21	0	2	2	2
7	g23	1	2	2	2
8	b4	0	3	2	1
9	b13	0	2	2	2
10	s8	1	3	2	3
11	b5	0	2	4	3
12	s4	1	3	5	2
13	b2	0	4	4	3
14	k27	1	2	2	2
15	b18	0	4	4	3
16	b29	0	2	2	2
17	b20	0	2	2	1
18	k1	1	5	5	4
19	b6	0	2	2	2

Radiograph Number	Type	Nodule Existence in Reality	Rating(1-5) For Medical Monitor	Rating(1-5) For Tablet	Rating(1-5) For Large Screen TV
20	k22	1	3	4	4
21	b15	0	1	2	2
22	k28	1	3	4	3
23	b26	0	2	2	2
24	b27	0	1	2	1
25	k16	1	4	1	4
26	k25	1	3	4	3
27	g17	1	2	4	4
28	b12	0	2	2	2
29	s7	1	3	2	2
30	b24	0	2	2	2
31	b7	0	1	2	1
32	b1	0	1	3	2
33	b8	0	4	1	4
34	b28	0	3	3	2
35	b10	0	4	2	2
36	k20	1	5	5	5
37	b22	0	4	2	1
38	k3	1	4	5	5
39	b11	0	1	2	2
40	b9	0	2	1	2
41	k4	1	4	3	3
42	k15	1	3	2	3

Radiograph Number	Type	Nodule Existence in Reality	Rating(1-5) For Medical Monitor	Rating(1-5) For Tablet	Rating(1-5) For Large Screen TV
43	k14	1	3	3	3
44	s5	1	2	2	2
45	b19	0	2	1	3
46	b17	0	2	1	1
47	b14	0	1	1	2
48	b3	0	2	2	2
49	b25	0	2	2	3
50	g27	1	2	2	3
51	g16	1	3	4	4
52	k12	1	4	4	4
53	b16	0	3	2	2
54	s2	1	3	3	2
55	g21	1	2	2	2
56	k2	1	4	5	3
57	b29	0	2	2	2
58	b23	0	2	1	1
59	s3	1	3	4	4
60	s6	1	1	3	2

Table 4.9
Results from Radiologist 3's evaluation on Medical Monitor; Tablet and Large Screen TV

Display Devices	Medical Monitor	Medical Monitor	Tablet	Tablet	Large Screen TV	Large Screen TV
Type of cases	P Cases	N cases	P Cases	N cases	P Cases	N cases
1 (definitely no nodule present)	1	6	2	5	1	6
2 (probably no nodule present)	6	16	10	20	8	17
3 (possibly nodule present)	13	3	3	3	10	6
4 (probably nodule present)	7	4	9	2	8	1
5 (definitely nodule present)	3	1	6	0	3	0
TOTAL	30	30	30	30	30	30

Table 4.10
Statistical Comparison of Radiologist 1's evaluations on Medical Monitors, Tablet and Large Screen TV

Display Devices	Medical Monitor	Tablet	Large Screen TV
Number of Cases	60	60	60
Number Correct	45	40	40
Accuracy	60%	66.7%	66.7%
Sensitivity	63.3%	63.3%	56.7%
Specificity	56.7%	70%	76.7%
Positive Cases Missed	11	11	13
Negative Cases Missed	13	9	7
Fitted ROC Area	0.747	0.699	0.759
Empiric ROC Area	0.721	0.691	0.744

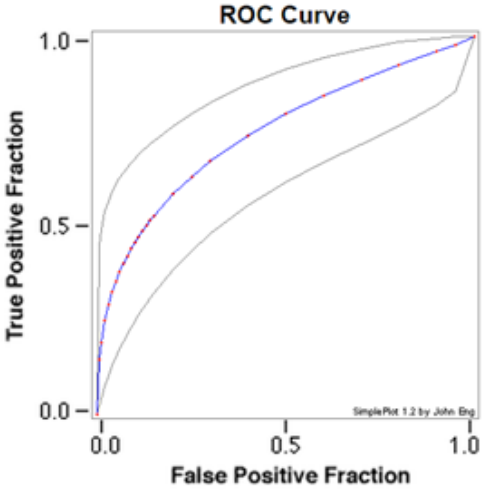


Figure 4.1 ROC Curve for Radiologist 1 with Medical Monitor

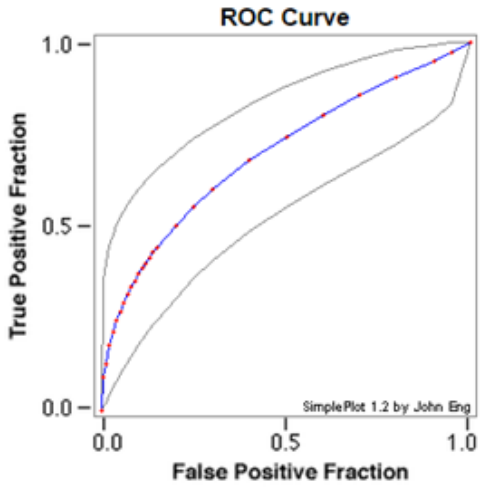


Figure 4.2 ROC Curve for Radiologist 1 with Tablet

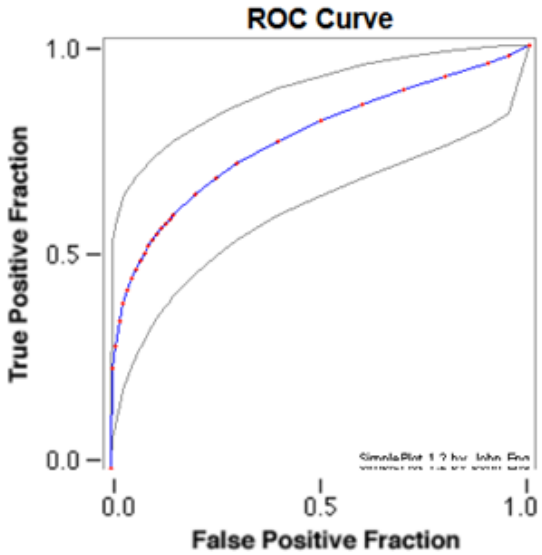


Figure 4.3 ROC Curve for Radiologist 1 with Large Screen TV

Table 4.11
Statistical Comparison of Radiologist 2's evaluations on Medical Monitors, Tablet and Large Screen TV

Display Devices	Medical Monitor	Tablet	Large Screen TV
Number of Cases	60	60	60
Number Correct	36	46	39
Accuracy	60%	76.7%	65%
Sensitivity	80%	60%	56.7% %
Specificity	40%	93.3%	73.3%
Positive Cases Missed	6	12	13
Negative Cases Missed	18	2	8
Fitted ROC Area	0.739	0.772	0.779
Empiric ROC Area	0.706	0.774	0.759

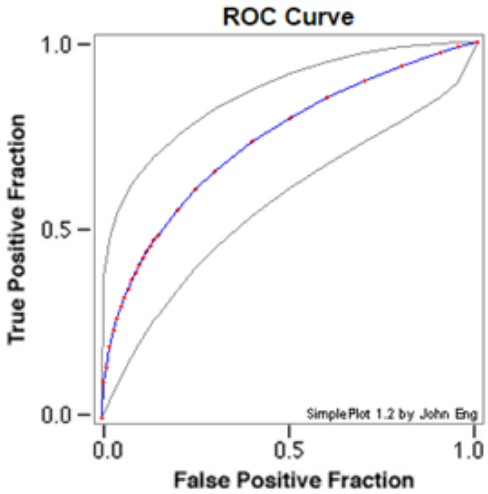


Figure 4.4 ROC Curve for Radiologist 2 with Medical Monitor

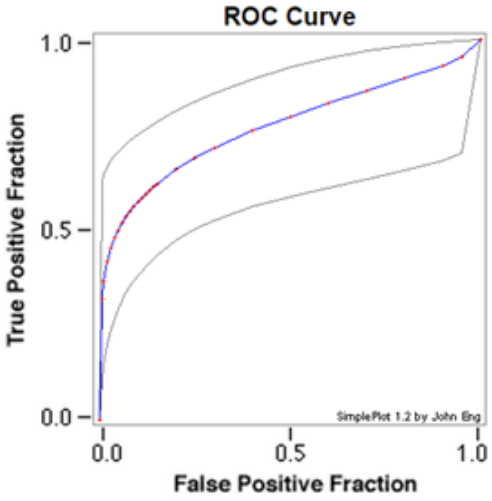


Figure 4.5 ROC Curve for Radiologist 2 with Tablet

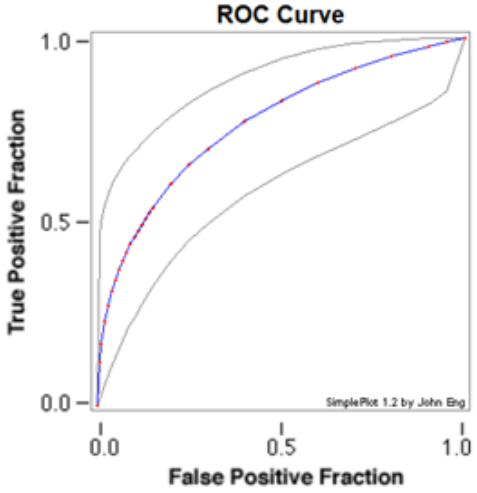


Figure 4.6 ROC Curve for Radiologist 2 with Large Screen TV

Table 4.12
Statistical Comparison of Radiologist 3's evaluations on Medical Monitors, Tablet and Large Screen TV

Display Devices	Medical Monitor	Tablet	Large Screen TV
Number of Cases	60	60	60
Number Correct	45	43	44
Accuracy	75%	71.7%	73.3%
Sensitivity	76.7%	60%	70%
Specificity	73.3%	83.3%	76.7%
Positive Cases Missed	7	12	9
Negative Cases Missed	8	5	7
Fitted ROC Area	0.759	0.776	0.811
Empiric ROC Area	0.744	0.751	0.783

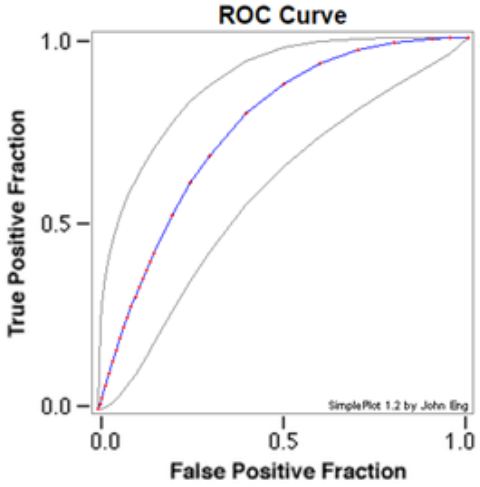


Figure 4.7 ROC Curve for Radiologist 3 with Medical Monitor

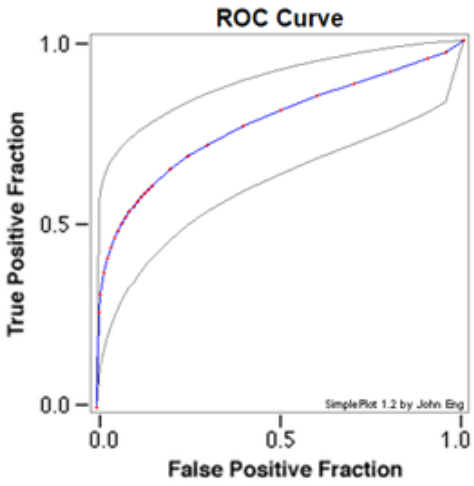


Figure 4.8 ROC Curve for Radiologist 3 with Tablet

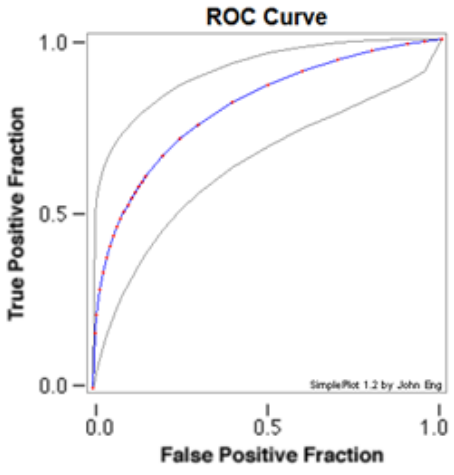


Figure 4.9 ROC Curve for Radiologist 3 with Large Screen TV

Table 4.13
Diagnostic performance (Area under the ROC Curve) of three observers for the diagnosis of nodule in chest X-ray radiographs on Medical Monitor, Tablet, and Large Screen TV through DBM-MRMC software

Area under the ROC curve			
Observer	Medical Monitor	Tablet	Large Screen TV
Radiologist 1	0.689	0.691	0.744
Radiologist 2	0.715	0.773	0.758
Radiologist 3	0.744	0.750	0.783
Overall	0.716	0.738	0.762

Table 4.14
95 confidence intervals and hypothesis tests (H0: difference = 0) for treatment AUC differences

Display systems comparison	Difference	Standard Error	P value	95 CI
Medical Monitor-Tablet	-0.0220	0.03137	0.4851	(-0.08479,0.04071)
Medical Monitor-Large Screen TV	-0.0457	0.03137	0.1500	(-0.10849,0.01701)
Tablet-Large Screen TV	-0.0237	0.03137	0.4528	(-0.08646,0.03905)

Table 4.15
Intra-observer agreement for diagnosis of nodule on chest x-ray on the Tablet and Medical monitor

Observer	Weighted kappa Tablet with Medical monitor	P value	Standard error
Radiologist 1	0.435	0.001	0.084
Radiologist 2	0.363	0.001	0.070
Radiologist 3	0.406	0.001	0.079
Overall	0.401	0.001	0.077

Table 4.16
Intra-observer agreement for diagnosis of nodule on chest x-ray on the Large Screen TV and Medical monitor

Observer	Weighted kappa Large Screen TV with Medical monitor	P value	Standard error
Radiologist 1	0.384	0.001	0.080
Radiologist 2	0.402	0.001	0.072
Radiologist 3	0.427	0.001	0.080
Overall	0.404	0.001	0.0773

Table 4.17
Total time-consuming and mean time

Time Table	Radiologist 1	Radiologist 2	Radiologist 3	Mean Time(s)	Mean Time(min)
Time-consuming with Medical Monitor	1031	1403	1316	1250	20 min 50 s
Time-consuming with Tablet	837	964	1046	896	14 min 56 s
Time-consuming with Large Screen TV	952	886	781	873	14 min 33 s
Mean Time(s)	940	1084	1047	-	-
Mean Time (min)	15 min 40 s	18 min 4 s	17 min 27 s	-	-

5. DISCUSSION AND CONCRETE RESULTS

Due to what mentioned in the previous chapter, the results of data analysis are given separately in the form of tables and figures.

The mean time of reviewing 60 chest x-ray radiographs by three radiologists with the Medical Monitor was 20 min 50 seconds, and with Tablet was 14min 56 seconds. It is obvious that the time-consuming for reviewing radiographs with the Tablet is less than the Medical Monitor. On the other hand,the mean time of reviewing radiographs with large screen TV was 14 min 33 seconds (Table 4.18). Therefore, the time-consuming for reviewing radiographs with Large screen TV in comparison with Medical Monitor and Tablet was more less in any two cases. It is clear that in the case of tablet and Large screen TV, there is time saving in diagnosis process in comparison with Medical Monitor.

Through Graphpod software, the weighted kappa values showed fair-to-moderate intra-observer agreement between the Tablet and Medical Monitor(Table 4.15). It was also shown fair-to-moderate intra-observer between the Medical Monitor and Large screen TV(Table 4.16). Furthermore, the weighted kappa value of more than 0.4 has clinical acceptance. The overall calculated intra-observer agreements between Medical monitor and Tablet was 0.401 whereas it was 0.404 between Medical monitor and Large screen TV. Therefore,in total the calculated kappa values are clinically acceptable except for radiologist 2 in calculation weighted kappa between tablet and medical monitor and for radiologist 1 in evaluation weighted kappa between large screen Tv and medical monitor.

Through JROCFIT 1.02.2 software tool [27],sensitivity and specifity values for diagnosis of nodule on radiographs in three types of display systems are given(Table 4.10-4.12). The all ROC curves were plotted by these data through this software.

Through analysis of Variance (ANOVA) software tool [17] , the area under the receiver operating characteristic (ROC) curves for the three types of displaying devices were compared by Dorfman-Berbaum-Metz method, using Rscore and semi-parametric estimation of ROC indices with DBM-MRMC software [26](Table 4.13). For all viewers, the calculated AUC values for the Tablet were higher than the calculated value for the Medical Monitor. It shows there is statistically no differences between the tablet and Medical Monitor. The calculated AUC values for the large screen TV also were much more higher than the calculated value for the Medical Monitor(Table 4.13). In the better explanation; the calculated overall AUC value with Large screen TV by three radiologists is 0.762 which is higher than its values with tablet and Medical monitor which are 0.738 and 0.716 respectively. Therefore, it is obvious that between the Medical Monitor and large screen TV, there is also statistically no difference and even the performance of Large screen TV in diagnosis chest x-ray is better than Tablet due to the calculated AUC values.

The comparison of the ROC results explained that with 95% confidence intervals the diagnosis performance of the Tablet and Large screen TV was similar to that of the Medical Monitor for analysis the nodule on radiographs (Table 14.13). Furthermore, the spent time for reviewing radiographs with Tablet and large screen TV was less than Medical monitor, and this causes to save time in evaluation process. Finally, it is possible to implement the Large screen TV and Tablet as a PACS monitor for medical diagnosis purposes without sacrificing any diagnostic value and time saving.

As a teleradiologic tool, tablet device has been introduced for mobile teleradiology via wireless networks. Although the data set was set via portable USB in tablet, it is possible to download the images via wireless networks. The outcomes of this study showed that the tablet can be introduced as a good hand-held teleradiologic tool because of its high performance without sacrificing any diagnostic value with time-saving property. It also has several remarkable advantages, including the relatively large display, its slim profile, light weight, and internet connectivity via Wi-Fi and 3G which all these producing enhanced portability. Furthermore, several DICOM viewer and wireless communication software are available for up-to-date android-based and IOS-

based tablet devices in order to receive and to operate DICOM image data. All these softwares can support the medical images viewing and evaluation in Tablet devices.

The survey form is prepared in order to get feedback of radiologist which the results are shown in Table 5.1-5.3. The quality of image, response time and user friendly items are investigated. The higher score indicates to have the better performance like score 5 for the quality of image which means the best quality and the score 1 which means the worst quality of image.

Table 5.1
Radiologist 1 point of view

Display systems	The quality of image(1-5)	response time(1-5)	User Friendly(1-5)
Medical Monitor	5	5	5
Tablet	3	2	2
Large Screen TV	4	3	3

Table 5.2
Radiologist 2 point of view

Display systems	The quality of image(1-5)	response time(1-5)	User Friendly(1-5)
Medical Monitor	5	5	4
Tablet	3	4	5
Large Screen TV	3	3	3

Table 5.3
Radiologist 3 point of view

Display systems	The quality of image(1-5)	response time(1-5)	User Friendly(1-5)
Medical Monitor	5	5	5
Tablet	4	4	3
Large Screen TV	3	3	4

According to radiologists' point of view, they seem to believe that medical grade monitor has the best image quality and the minimum response time. Also according to their opinion, unlike tablet and large screen TV, medical monitors seemed to be less eye-tiring. However, results of this study reveal that there are no significant differences between performance of tablet and large screen TV in comparison with the medical monitor. In addition, the response time in viewing images was much faster in tablet and large screen TV. This fact indicates that there is a mistaken prejudgment regarding performance quality of tablet and large screen TV.

The radiologists‘ assessment on the ease of usage of the tablet indicates that it is more user friendly since tablet is one of the best hand-held Teleradiology devices. As for the large screen TV, the lesion detectability was easier due to largeness of the screen which eliminated the need for zoom.

The fact remains obvious that all radiologists are fond of medical monitors and trust in the precision of diagnosis based on these medical monitors because they are used to using these devices for a long time period and mistakenly believe that the diagnosis of nodules in large screen TV and tablet was not done properly.

REFERENCES

1. Dunn, S. M. and M. L. Kantor, "Digital radiology facts and fictions", *The Journal of the American Dental Association*, Vol. 124, No. 12, pp. 39–47, 1993.

2. Weatherburn, G., S. Bryan, A. Nicholas and R. Cocks, "The effect of a picture archiving and communications system (PACS) on diagnostic performance in the accident and emergency department", *Journal of accident & emergency medicine*, Vol. 17, No. 3, pp. 180–184, 2000.

3. Chen, S.-K., "Integration of the digital imaging and communications in medicine standard into an oral and maxillofacial image management and communication system", *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, Vol. 91, No. 2, pp. 235–238, 2001.

4. Haak, R., M. Wicht, M. Hellmich, G. Nowak and M. Noack, "Influence of room lighting on grey-scale perception with a CRT-and a TFT monitor display", *Dentomaxillofacial Radiology*, Vol. 31, No. 3, pp. 193–197, 2002.

5. Antonuk, L. E., "a-Si: H TFT-based active matrix flat-panel imagers for medical x-ray applications", *Thin film transistors, materials and processes*, Vol. 1, pp. 395–484, 2003.

6. Geijer, H., M. Geijer, L. Forsberg, S. Kheddache and P. Sund, "Comparison of color LCD and medical-grade monochrome LCD displays in diagnostic radiology", *Journal of digital imaging*, Vol. 20, No. 2, pp. 114–121, 2007.

7. Mariani, C., A. Tronchi, L. Oncini, O. Pirani and R. Murri, "Analysis of the X-ray work flow in two diagnostic imaging departments with and without a RIS/PACS system", *Journal of digital imaging*, Vol. 19, No. 1, pp. 18–28, 2006.

8. Mc Laughlin, P., S. O. Neill, N. Fanning, A. M. Mc Garrigle, O. J. Connor, G. Wyse and M. M. Maher, "Emergency CT brain: preliminary interpretation with a tablet device: image quality and diagnostic performance of the Apple iPad", *Emergency radiology*, Vol. 19, No. 2, pp. 127–133, 2012.

9. Shiraishi, J., L. L. Pesce, C. E. Metz and K. Doi, "Experimental Design and Data Analysis in Receiver Operating Characteristic Studies: Lessons Learned from Reports in Radiology from 1997 to 2006 1", *Radiology*, Vol. 253, No. 3, pp. 822–830, 2009.

10. Gönen, M., *Analyzing receiver operating characteristic curves with sas*, SAS Institute, 2007.

11. Cook, N. R., "Use and misuse of the receiver operating characteristic curve in risk prediction", *Circulation*, Vol. 115, No. 7, pp. 928–935, 2007.

12. Zou, K. H., A. J. O'Malley and L. Mauri, "Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models", *Circulation*, Vol. 115, No. 5, pp. 654–657, 2007.

13. Streiner, D. L. and J. Cairney, "What's under the ROC? An introduction to receiver operating characteristics curves", *Canadian Journal of Psychiatry*, Vol. 52, No. 2, p. 121, 2007.

14. DeLong, E. R., D. M. DeLong and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach", *Biometrics*, pp. 837–845, 1988.

15. Pepe, M. S., *The statistical evaluation of medical tests for classification and prediction*, Oxford University Press, 2003.

16. Buls, N., W. Shabana, P. Verbeek, P. Pevenage and J. De Mey, "Influence of display quality on radiologists' performance in the detection of lung nodules on radiographs", *The British journal of radiology*, 2014.

17. Miller Jr, R. G., *Beyond ANOVA: basics of applied statistics*, CRC Press, 1997.

18. Gueorguieva, R. and J. H. Krystal, "Move over anova: Progress in analyzing repeated-measures data andits reflection in papers published in the archives of general psychiatry", *Archives of general psychiatry*, Vol. 61, No. 3, pp. 310–317, 2004.

19. Crowder, M. J., "Beta-binomial anova for proportions", *Applied Statistics*, pp. 34–37, 1978.

20. Viera, A. J., J. M. Garrett *et al.*, "Understanding interobserver agreement: the kappa statistic", *Fam Med*, Vol. 37, No. 5, pp. 360–363, 2005.

21. Cohen, J., "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.", *Psychological bulletin*, Vol. 70, No. 4, p. 213, 1968.

22. Wade, C. and P. Brennan, "Assessment of monitor conditions for the display of radiological diagnostic images and ambient lighting", *The British journal of radiology*, 2014.

23. Samei, E., A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M. J. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson *et al.*, "Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report", *Medical physics*, Vol. 32, No. 4, pp. 1205–1225, 2005.

24. Kayabasi, S. H., "Design Of A 32 Color TFT LCD PACS Monitor and Its Clinical Evaluation Through ROC Analysis", *Bogazici University Biomedical Engineering Institute*, , No. 335, 2011.

25. Tourassi, G., S. Voisin, V. Paquit and E. Krupinski, "Investigating the link between radiologists' gaze, diagnostic decision, and image content", *Journal of the American Medical Informatics Association*, Vol. 20, No. 6, pp. 1067–1075, 2013.

26. Berbaum, K., C. Metz, L. Pesce *et al.*, "DBM MRMC user's guide", , 2006.

27. Hirt, M. and M. Lagergren, *Oracle JRockit: The Definitive Guide*, Packt Publishing Ltd, 2010.