ASSESSING A VIRTUAL REALITY DISPLAY SYSTEM FOR TELERADIOLOGY PURPOSES USING THE RECEIVER OPERATING CURVE METHODOLOGY

by

Eylem Kocaturk

B.S., Teaching Physics, Boğaziçi University, 1999

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Physics

Boğaziçi University

2018

ASSESSING A VIRTUAL REALITY DISPLAY SYSTEM FOR TELERADIOLOGY PURPOSES USING THE RECEIVER OPERATING CURVE METHODOLOGY

APPROVED BY:

Prof. Dr. Levent Kurnaz (Thesis Supervisor)	.....
Assoc. Prof. Dr. Albert Guvenis (Thesis Co-supervisor)	.....
Prof. Dr. Naci Inci	.....
Assist. Prof. Dr. Can Kozcaz	.....
Assoc. Prof. Dr. Neslihan Tasdelen	.....

DATE OF APPROVAL: 12.01.2018

ACKNOWLEDGEMENTS

Firstly, I would like to acknowledge my thesis advisors Prof.Dr.Mehmet Burcin Unlu, Prof.Dr. Levent Kurnaz and Dr. Albert Guvenis for providing me with opportunities and support.

I would also like to express my gratitude to all that have welcomed and supported me at Yeditepe University Hospital. I am fully grateful to them who evaluated the data set in two types of display devices in their tight and stressful schedule. Their guidance has been invaluable and I would like to especially thank Dr. Neslihan Tasdelen, Dr.Aysegul Sarsilmaz, Dr. Basar Sarikaya, Dr. Melih Topcuoglu and Dr. Emrah Karatay.

The involvement of Dr. Neslihan Tasdelen has also been instrumental for this work and I am grateful for her help, feedback, and advice. I would also like to thank the MSI Inc. Turkey to provide me the high-tech laptop.

I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. My thanks and gratitude deep from my heart goes to my husband, without his support and motivation, I would not dare to start the thesis study after the long break.

Finally, I would like to thank specially to my son, Umut Burak Kocaturk, for his understanding while working on this thesis.

ABSTRACT

ASSESSING A VIRTUAL REALITY DISPLAY SYSTEM FOR TELERADIOLOGY PURPOSES USING THE RECEIVER OPERATING CURVE METHODOLOGY

In a radiology reading room, multiple flat displays are used for the comparison aspect. This approach increases the overall expense and also requires more space. So, virtual reality techniques can be applied to teleradiology. By using VR devices, radiologists may complete diagnostic tasks outside of hospital settings.

In this study, 100 lung X-ray images with definite findings were gotten from the database approved by the radiologists. Five radiologists from Yeditepe University Hospital assessed these images both on a medical grade LCD monitor and on the Oculus Rift CV1. The receiver-operating characteristic (ROC) curves for all display-reader cases were statistically compared to check observers' performances. AUC for virtual reality head mounted display for reader 1, 2, 3, 4,5 and overall were calculated as 0.931, 0.865, 0.805, 0.775, 0.728 and 0.814 respectively. AUC for medical monitor were 0.948, 0.915, 0.897, 0.771, 0.769 and 0.861.

Sensitivity and specificity of p values from McNemar test were 0,37 and 0.90 respectively. Since, the p values are greater than 0.05, this is a good indicator of no difference between two diagnostic techniques. Weighted kappa results for reader 1, 2, 3, 4, 5 and overall were 0.656, 0.782, 0.557, 0.383, 0.224 and 0.542 respectively. So, the overall weighted kappa value showed the moderate agreement between two diagnostic techniques.So, Oculus Rift CV1 can be used as a diagnostic display by the radiologists.

ÖZET

SANAL GERÇEKLİK EKRAN SİSTEMİNİN TELERADYOLOJİDE KULLANIMININ İŞLEM KARAKTERİSTİK EĞRİSİ ANALİZİYLE DEĞERLENDİRİLMESİ

Bir radyoloji okuma odasında, karşılaştırma yapabilmeleri için çok sayıda düz ekranlar kullanılır. Bu, genel masrafı arttırır ve ayrıca daha fazla alan gerektirir. Dolayısıyla sanal gerçeklik teknikleri teleradyolojiye uygulanabilir. Radyologlar, VR cihazlarını kullanarak teşhis görevlerini hastane dışında tamamlayabilirler.

Bu çalışmada radyologlar tarafından onaylanan veri tabanından elde edilen ve kesin bulguları olan 100 akciğer grafisi kullanılmıştır. Yeditepe Üniversitesi ne bağlı hastanede çalışan beş radyolog, bu görüntüleri tıbbi kalitede bir LCD monitörde ve Oculus Rift CV1'de değerlendirdi. Tüm görüntü okuyucu durumlarda, alıcı-işletim karakteristiği (ROC) eğrileri gözlemcilerin performansını istatistiksel olarak karşılaştırdı. Okuyucu 1, 2, 3, 4, 5 ve toplam için sanal gerçeklik baş monteli ekran için AUC sırasıyla 0.931, 0.865, 0.805, 0.775 , 0.728 ve 0.814 olarak hesaplandı. Tıbbi monitör için AUC, 0.948, 0.915, 0.897, 0.771, 0.769 ve 0.861 idi.

McNemar testinden, p değerlerinin duyarlılığı ve özgüllüğü sırasıyla 0,37 ve 0,90 idi. p değerleri 0.05'den büyük olduğundan, iki tanı tekniği arasında fark olmadığı görüldü. Okuyucular 1, 2, 3, 4, 5 için ağırlıklı kappa sonuçları sırasıyla 0.656, 0.782, 0.557, 0.383, 0.224 ve 0.542 idi. Bu nedenle, toplam ağırlıklı kappa değeri, iki tanı tekniği arasında orta derecede bir uyum gösterdi. Bu nedenle, Oculus Rift CV1, radyologlar tarafından bir teşhis ekranı olarak kullanılabilir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZET v
LIST OF FIGURES viii
LIST OF TABLES xi
LIST OF SYMBOLS xiii
LIST OF ACRONYMS/ABBREVIATIONS xiv
1. INTRODUCTION 1
2. BACKGROUND 3
2.1. X-Ray Physics 3
2.1.1. Conventional Projection Radiographic System 3
2.1.2. Photoelectric Absorption 5
2.1.3. Rayleigh Scattering 6
2.1.4. Compton Scattering 6
2.1.5. Pair Production 7
2.2. PACS (Picture Archiving and Communication System) 7
2.3. Virtual Reality Technology 9
2.3.1. Modern Virtual Reality Devices for Consumer Use 10
2.3.2. Applicaton of Virtual Reality in Radiology 10
2.3.3. Field of View 11
2.3.4. Portability and Cost 11
3. LESION DETECTION 13
3.1. Causes of Missed Lung Cancer 14
3.1.1. Observer Error 14
3.1.2. Lesion Characteristics 15
3.1.3. Technical Concerns 17
3.2. Approaches to Decrease the Occurrence of Missed Pulmonary Malignancies on Chest Radiography 17
3.2.1. Reducing Observer Error 17

3.2.2. Improving Tumor Detection Techniques 18
4. METHODOLOGY 20
4.1. ROC Analysis 20
4.2. McNemar's Test 27
4.3. Kappa Testing 28
5. PROTOTYPE DESIGN 30
5.1. Display 30
5.2. Software 31
5.3. System Design 32
5.4. Assessment of Display Performance 34
6. EXPERIMENTS AND RESULTS 37
6.1. Differential Diagnosis Experiment 37
6.1.1. Experiment Design 37
6.1.2. Experiment Results 39
6.1.3. Experiment Discussion 55
7. CONCLUSION 61
REFERENCES 62

LIST OF FIGURES

Figure 2.1. Bremsstrahlung and characteristic radiation for a tungsten anode through X-ray production at different kVp values. 4

Figure 2.2. Illustrative summary of X-ray and Gamma ray interactions.(A) represents unattenuated beam (B) represents Photoelectric Absorption (C) represents Rayleigh scattering D) represents Compton Scattering. 5

Figure 2.3. Virtual Reality spectrum. 9

Figure 2.4. The schematics represents the approximate field of view for Oculus Rift Consumer Version compared to a standard medical display from 50 cm away. 11

Figure 3.1. There is a nodular opacity on the left shown as an arrowhead. Nodular opacity shown as an arrow was not reported because it is considered as a nipple. 15

Figure 3.2. An oval opacity in the right upper lobe shown as an arrow was missed by the radiologist. 16

Figure 3.3. The representative figure of chest reading methodology. All lung zones (shown by arrows) are scanned. Stars show the “blind zones” and lines show the mediastinal lines and stripes. 18

Figure 4.1. The classification table (counts). 21

Figure 4.2. The fitted ROC curve (formed through the assumption of a binormal distribution). 23

Figure 4.3. The empirical ROC curve. Dots on the graph represent the discrete points 24

Figure 4.4. Four ROC curves. Curve A represents a perfect test (has an AUC of 1). ROC curves B and C can discriminate between those subjects with and without disease. Diagnostic performance of Test B (higher AUC) is better than test C. Curve D that is called as chance diagonal has an AUC of 0.5. 24

Figure 4.5. Although curve A and B have equal AUC, they are not identical. Test A has better sensitivity range however test B is better than test A in the high FPR range. 26

Figure 5.1. Oculus Rift Consumer Version infrared markers. 31

Figure 5.2. The system design layout with Oculus Rift HMD 2016. 33

Figure 5.3. TG18-QC test pattern. It is used for quantitative assessment of geometric distortions. 35

Figure 5.4. The TG18-CX test pattern for the assessment of display resolution and resolution uniformity. 36

Figure 6.1. Evaluation of lung X-ray images through Oculus Rift VR HMD. . 38

Figure 6.2. A sample X-ray lung image from the database seen through medical monitor. 40

Figure 6.3. Reader 1's evaluation of lung X-Ray Images on VR and MM. . . . 41

Figure 6.4. A: ROC curve of VR and B: ROC curve of MM for the Reader 1. 42

Figure 6.5. Reader 2's evaluation of lung X-Ray images on VR and MM. . . . 44

Figure 6.6. A: ROC curve of VR and B: ROC curve of MM for the Reader 2. 45

Figure 6.7. Reader 3's evaluation of lung X-Ray images on VR and MM. . . . 46

Figure 6.8. A: ROC curve of VR and B: ROC curve of MM for the Reader 3. 47

Figure 6.9. Reader 4's evaluation of lung X-Ray images on VR and MM. . . . 48

Figure 6.10. A: ROC curve of VR and B: ROC curve of MM for the Reader 4. 49

Figure 6.11. Reader 5's evaluation of lung X-Ray images on VR and MM. . . . 50

Figure 6.12. A: ROC curve of VR and B: ROC curve of MM for the Reader 5. 51

Figure 6.13. Sample missed case with MM. 57

Figure 6.14. Sample missed case with MM. 58

Figure 6.15. Sample missed case with VR. 59

Figure 6.16. Sample missed case with VR. 59

Figure 6.17. Sample missed case with both VR and MM. 60

LIST OF TABLES

Table 2.1. Minimum and recommended specifications for primary diagnostic display devices used in clinics 8

Table 4.1. Data summary of 2x2 matrix of McNemar Test 27

Table 4.2. The summary of Kappa values and the level of agreement 29

Table 5.1. The technical comparison of medical monitor and Oculus Rift CV1. 32

Table 6.1. Time consumed while assessing 25 X-ray lung nodules on VR and medical monitor (MM). 39

Table 6.2. Statistical comparison of Reader 1's evaluation on VR and MM . . 42

Table 6.3. Statistical comparison of Reader 2's evaluation on VR and MM . . 43

Table 6.4. Statistical comparison of Reader 3's evaluation on VR and MM . . 45

Table 6.5. Statistical comparison of Reader 4's evaluation on VR and MM . . 47

Table 6.6. Statistical comparison of Reader 5's evaluation on VR and MM . . 51

Table 6.7. Area Under Curve (AUC) for VR and MM 52

Table 6.8. Sensitivity and specificity for the diagnosis of X-ray lung nodules using VR and MM. 52

Table 6.9. Reviewers' (n=5) overall sensitivity for 50 nodule cases. 53

Table 6.10. Reviewers' (n=5) overall specificity for 50 non-nodule cases. 53

Table 6.11. Intra-observer agreement with VR and MM for all readers. 53

LIST OF SYMBOLS

E_{BE}	Electron Binding Energy
E_S	Scattered Photon Energy
E_0	Incident Photon Energy

LIST OF ACRONYMS/ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
AUC	Area Under Curve
CT	Computed Tomography
CV1	Consumer Version 1
DICOM	The Digital Imaging and Communication in Medicine
DM	Decision Matrix
FP	False Positives
FPR	False Positive Rate
GSDF	Grayscale Standard Display Function
HMD	Head Mounted Display
HVS	Human Visual System
IMU	Inertial Measuring Unit
IPD	Inter Pupillary Distance
JND	Just Noticible Difference
MeV	Mega electron Volt
MIPAV	Medical Imaging Processing Analysis and Visualization
MRI	Magnetic Resonance Imaging
MS	Matched Sample
MM	Medical Monitor
OLED	Organic Light Emitting Diode
PACS	Picture Archiving and Communication System
PET	Positron Emission Tomography
ROC	the Receiver Operating Characteristic
RADAR	Radio Signal Detecting and Ranging
TP	True Positives
VR	Virtual Reality
Z	Atomic Number

1. INTRODUCTION

In medical imaging procedures, diagnosis and treatment are based on the assessment of image data. In most cases, the human interpretation of data using visualization approaches and display devices is the final stage of image process.

Depending on the resolution, the cost of Medical Diagnostic Display devices ranges from $1.500 -$30.000. In order to diagnose and interpret medical images, radiologists spend their valuable time in the dark reading room that is called picture archiving and communication system (PACS) reading room.

The advancement in wearable technology aims to enhance the quality of life for humans. So, Virtual Reality (VR) devices can provide a solution to this issue faced by radiologists. By using VR devices, radiologists may complete diagnostic tasks outside of hospital settings. The modern Virtual Reality display has adequate resolution and image quality to identify tumor in Computed Tomography (CT) images [1].

In addition, the development of wide-scale implementations of easily accessed picture archive and communication systems (PACS) and the availability of wireless connectivity has given a rise to the existence of telemedicine applications.

The aptness of handheld devices including laptops, tablets, phones, and VR are gaining popularity in everyday consultation diagnosis of medical professionals. Radiologists can perform diagnostic tasks in a virtual environment rather than standard reading room by using VR technology. In some cases, clinicians may not have an access to these standard reading rooms. This can also lead to allow the collaboration between the radiologists. Additionally, VR technology decreases the investment of setting PACS rooms.

The advancement of mobile wearable technology and the importance of portable diagnostic displays for teleradiology was briefly mentioned above, the outline of further chapters in the thesis will be as follows:

In Chapter 2, X-Ray physics, a brief history of virtual reality and consumer virtual reality displays, benefits of modern virtual reality displays for radiological practice will be discussed.

In Chapter 3, difficulty in detection lung lesions on chest radiography, the reasons of missed lung cancer and some approaches to reduce the lung cancer will be mentioned.

In Chapter 4, the basic principles of statistical tests such as ROC analysis, McNemar Test and Kappa Test will be given for comparison of two diagnostic techniques.

In Chapter 5, the components and the specifications of the virtual reality and also the assessment of the display performance will also be discussed.

In Chapter 6, experiment design, evaluation of lung X-ray images by using Oculus Rift CV1 virtual reality head mounted display and Medical monitor will be presented. Statistical analysis of both diagnostic techniques will also be shown.

In Chapter 7, the summary of the experiment and the future applications of the virtual technology in teleradiology will be discussed.

2. BACKGROUND

2.1. X-Ray Physics

X-ray imaging is a projection based medical imaging method that is widely used in radiology. In conventional radiography which is also called projection radiography, the three dimensional volume of the human anatomy is projected onto a two dimensional imaging surface.

Theoretically, the projection radiograph characterizes the X-ray beam attenuation through the human body and it is weighted by integrated loss of beam energy because of scattering and absorption in the body.

Short exposure time, large field of view, low radiation exposure,low cost, excellent contrast and spatial resolution are some advantages of projection radiographic systems. Projection Radiography is used to detect some diseases such as pneumonia, heart disease, lung disease, bone fracture, cancer and vascular disease.

2.1.1. Conventional Projection Radiographic System

X-rays travels through the patient body as a short pulse beam. Ultimately, non-absorbed X-ray photons within the patient and scattered X-ray photons create an anatomic image on a sheet of film.

Nuclear medicine imaging is an essential tool for diagnostic radiology. Characteristic X rays can be produced by bombarding a metal target (i.e. tungsten) in an x-ray tube with high-energy electrons. During this interactions, the high energy electrons can eject inner-shell electrons of the target metal atoms. The vacancies are filled by electrons belong to higher energy shells that emit X-rays with sharply defined frequencies. These frequencies are characteristic of elements and they are associated with the atomic energy level differences for each element.Bremsstrahlung that means ”breaking

rays" in German is another mechanism to generate X-rays. In the Bremsstrahlung process, high energy electrons moving towards the target metal are slowed down or completely stopped by the electrostatic forces. According to the law of conservation of energy, this energy must be absorbed by the atom. The excessive energy then is radiated as X-rays Figure 2.1 [2].

In conventional radiography, when the incident beam passes through human body, X-ray photons are absorbed differently. Each tissue type causes different X-ray attenuation ratio and this difference in X-ray penetration represents the tissue contrast in final medical image.

In the absence of contrast media, Compton effect is responsible for the X-ray contrast for soft tissue whereas in addition to the Compton effect, the photoelectric effect plays an essential role for X-ray contrast of bone tissue. Also, contrast agents can be used to enhance soft tissue contrast when necessary.

Medical X-ray imaging based on four main X-ray interactions which are photoelectric absorption, Rayleigh scattering, Compton scattering, and pair production Figure 2.2 [3].

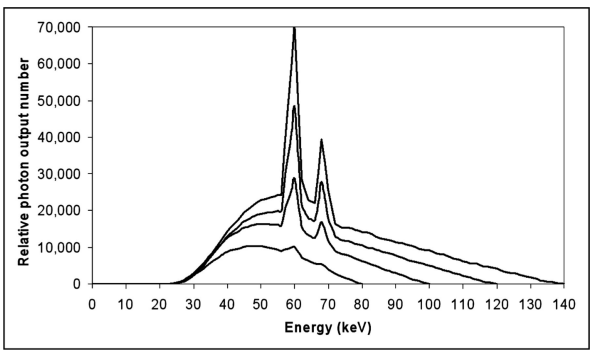


Figure 2.1. Bremsstrahlung and characteristic radiation for a tungsten anode through X-ray production at different kVp values.

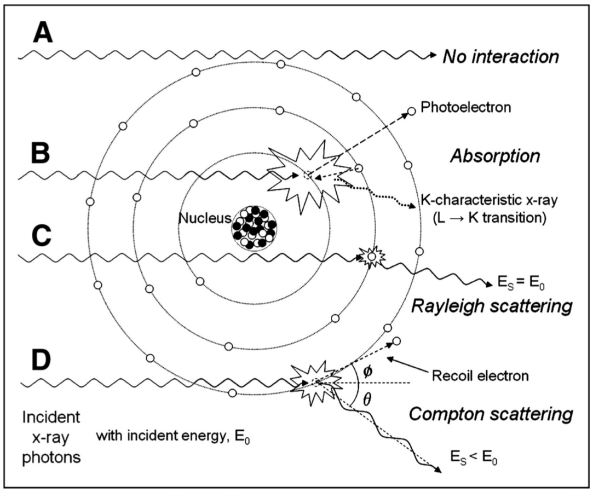


Figure 2.2. Illustrative summary of X-ray and Gamma ray interactions.(A) represents unattenuated beam (B) represents Photoelectric Absorption (C) represents Rayleigh scattering D) represents Compton Scattering.

.

2.1.2. Photoelectric Absorption

Photoelectric absorption occurs when a low energy photon interacts with the inner shell electron in the atom and it removes the electron from its shell. In X-ray imaging, the incident X-ray photon transfers its energy to the inner shell electron and then the electron ejects from its shell (usually the K shell) with a kinetic energy equals to the difference of the incident photon energy and electron binding energy. The absorption of X-ray photons cause increasing in attenuation. However, photoelectric interaction can appear only if the incident X-ray photon energy is higher than the binding energy of the electron. As the incident photon energy is higher than the electron shell binding energy, photoelectric absorption rate decreases at a rate proportional to $1/E^3$ where E is the incident photon energy. After photoelectric interaction, the atom ionizes and forms a free electron and a positively charged atom. The ejected electrons can cause further ionization if the kinetic energy of the ejected photoelectron is sufficient. Hence, the incident photon energy is accumulated locally for photoelectric absorption in tissues.

The probability of photoelectric absorption (p) can be calculated by

$$p = \frac{Z^3}{E^3} \tag{2.1}$$

where Z is the atomic number of the interacting atom and E is the incident photon energy. In other words, when X-ray beam passes through a high atomic number element the occurrence rate of photoelectric absorption increases. Therefore, X-ray detectors, protection devices and contrast agents are usually made of elements with high atomic numbers such as lead,iodine, gadolinium respectively.

2.1.3. Rayleigh Scattering

In Rayleigh scattering, an incident X-ray photon can interact with an electron and deflected without losing its energy. This phenomena is called as elastic scattering. During this process, the interacted electron energy temporarily increases without leaving from its orbital. In medical imaging, most X-rays are scattered forward with a small angle through this mechanism. The atom cannot experience recoil due to the momentum conservation law because the interacted electron stays in its orbit. If the atomic number of the atoms interacted with X-ray increases the possibility of Rayleigh scattering also increases. If the X-ray beam energy decreases then the possibility of Rayleigh scattering decreases.

2.1.4. Compton Scattering

Compton scattering phenomena occurs when an inelastic collision occurs between an X-ray photon of energy E_0 that is much greater than the binding energy of an atomic electron (outer shell electron). Throughout this interaction, X-ray photons partially transfer its kinetic energy to the electron which is removed from the atom at a certain angle whereas X-ray photon recoils. The remainder of incident X-ray photon energy, E_S, is transferred to a scattered X-ray photon. The scattered photon follows a trajectory of angle Θ relative to the trajectory of the incident photon. The scattered

photon energy relative to the incident photon energy for a certain scattering angle is given by the equation 2.2 that is named as the Klein-Nishina equation.

$$\frac{E_s}{E_0} = \frac{1}{1 + \frac{E_0}{511keV}(1 - cos\theta)} \tag{2.2}$$

where E_s is the scattered photon energy, E_0 is the incident photon energy, 511 keV is the energy equal to the rest mass of the electron. Qualitatively, this equation shows that the scattered X-ray energy depends on the scattering angle and as the scattering angle increases the scattered X-ray energy gets smaller. When the incident photon energy increases this effect is also amplified.

2.1.5. Pair Production

If the incident X-ray or Gamma-ray photon has energy greater than 1.02 MeV (the rest mass energy equivalent of 2 electrons) and interacts with the nucleus of an atom the photon energy transform into an electron and positron. The probability of pair production increases with proportional to the square of atomic number of interacted nucleus. Therefore, the pair production is not a dominant process through X-ray beam-soft tissue interaction and not likely to occur in the energy ranges used in diagnostic radiology.

2.2. PACS (Picture Archiving and Communication System)

A picture archiving and communication system (PACS) is used to replace the roles of conventional radiological films. PACS obtains, stores, transfers, and displays images digitally. Modern PACS includes many components such as a controller, image acquisition components, a server, an archiving system for database and a primary network frame to connect them. PACS uses industrial standard formats such as DICOM and Health Language Level 7 (HL7). This feature is essential to be compatible with new imaging modalities. The digital imaging and communication in medicine (DICOM) grayscale standard display function (GSDF) describes the luminance response

of a display such that an observer's perception of image contrast is consistent throughout the pixel value range of a displayed image. The response of the Barten model of the human visual system (HVS) to contrast stimuli has been used to develop the digital imaging and communication in medicine (DICOM) grayscale standard display function (GSDF) [4] for grayscale medical images .

The grayscale calibration is mandatory for electronic medical monitors because the calibration process aims to guarantee that every observer may have perceived medical images with the same contrast appearance on all displays, which does not depend on the display luminance range. The Barten model measures the sensitivity of the HVS to a low contrast, sinusoidal luminance signal presented on uniform luminance backgrounds. The DICOM grayscale display function is often plotted graphically. In the graph, the luminance versus just noticeable difference (JND) index (j), where the JND index values range from 0 to 1023 .The DICOM grayscale standard display function provides a relationship between luminance and monitor input values such that the slope of the curve at any point. Table 2.1 shows the minimum and recommended specification of PACS for diagnostic displays used for clinical image interpretation at The Royal College of Radiologists.

Table 2.1. Minimum and recommended specifications for primary diagnostic display devices used in clinics.

	Minimum	**Recommended**
Screen Resolution	≥1280 x 1024	≥1500 x 2000
Screen Size	≥42 cm	≥50 cm
Maximum Luminance	≥170 cd/ m^2	≥500 cd/ m^2
Luminance Resolution	≥ 250:1	≥ 500:1
Grayscale Contrast	Within %10 GSDF	Calibrated to GSDF
Grayscale Bit Depth	8 bit grayscale	10 bit grayscale

2.3. Virtual Reality Technology

Virtual reality is an immersive environment. A person interacts with this environment artificially. Figure 2.3 shows the virtual reality spectrum. The difference between virtual reality and augmented reality is that in virtual reality, users are totally engaged with a virtual simulated environment [5].

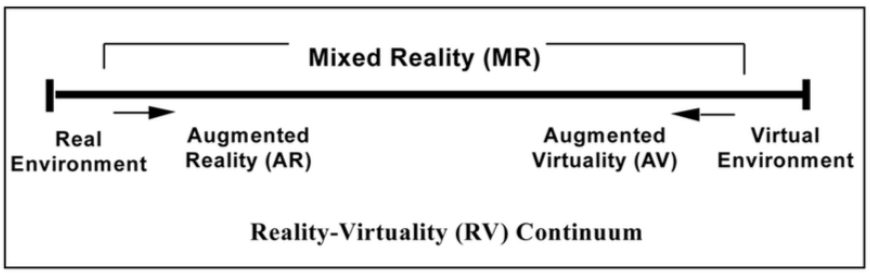


Figure 2.3. Virtual Reality spectrum.

Virtual reality (VR) technology has been developed significantly after the experiments with the "Sword of Damocles" by Sutherland [6]. Primarily, it was developed restrictedly for multi-million dollar military projects only. In the late 1980s, consumer-oriented VR devices have been emerged but with limited practical applications for consumer market. Both head and hand tracking systems, resolution of the head display monitor, auxiliary interfaces, latency, and video graphics rendering speed are very essential for satisfactory VR experiences. On the other hand, the VR system portability, cost, the comfort of the head set and convenience of the overall system are important for an effective consumer-oriented device. Absences in all or some of these areas limit the usage of such system.

In the past, many virtual reality systems have been designed and developed for considering a wide variety of potential medical applications. Pieper *et al.* [7] introduced the first stereo head-tracked display for minimally invasive surgery in 1994. Calvano *et al.* [8] used the first head-mounted display to perform endoscopic fetal surgery under real time three dimensional imaging. In 1998, virtual reality environment was used for

virtual bronchoscopy and virtual angioscopy by Krapichler [9]. The implementation of virtual reality devices for medical applications was limited due to the high cost and insufficient interface systems.

2.3.1. Modern Virtual Reality Devices for Consumer Use

Recently, the demand for low cost consumer virtual reality systems has been increased. Recent tehnological developments such as smart phone development, real time high resolution video rendering in video games arise the public attention. The development of the low cost and high resolution displays and motion sensors has played a crucial role to bring new VR devices such as the Oculus Rift, introduced by Oculus VR (Irvine, CA) [10] into the consumer market in late 2012. The Oculus Inc. was acquired by Facebook, inc. in 2014 for 2 billion USD.

Mostly, it seems that current application for the new consumer-oriented VR head mounted displays is for the video game and entertainment industry. Overcoming limitations in weight, accuracy, and cost increase the practical use of head-mounted displays. Now, the current model of the Oculus Rift HMD that is released in late 2016 weighs 470 g and costs 600 USD.

2.3.2. Applicaton of Virtual Reality in Radiology

Virtual reality is being used in the medical field for some purposes such as medical training, scientific animations, surgery planning, during-operation assistance. Virtual reality has applications also in dentistry, oto-laryngology, therapy for phobias and fears, minimally invasive general surgery, balance dysfunction, training for patient interviews and endoscopy.

Currently, radiologists spend a considerable amount of time while evaluating medical images from diagnostic medical displays in picture archiving and communication system (PACS) reading rooms. The implementation of PACS reading rooms has some advantages such as improving efficiency and decreasing the use of space instead of using

film [11]. Conversely, according to the surveys of radiologists presented by Rumreich & Johnson [12], radiologists suffer from poor ambient lighting, lack of space due to large multi-monitor workstations. Likewise, Siegel *et al.* [13] declared the adverse effects of poor lighting and multiple monitors on radiologist productivity. These concerns and existing problems confronted by radiologists practice in PACS reading rooms may be solved by the arrival of modern head-mounted VR devices.

2.3.3. Field of View

Virtual reality displays have a large field of view as demonstrated in Figure 2.4. Large displays have advantages for 3D and 2D visual scanning.

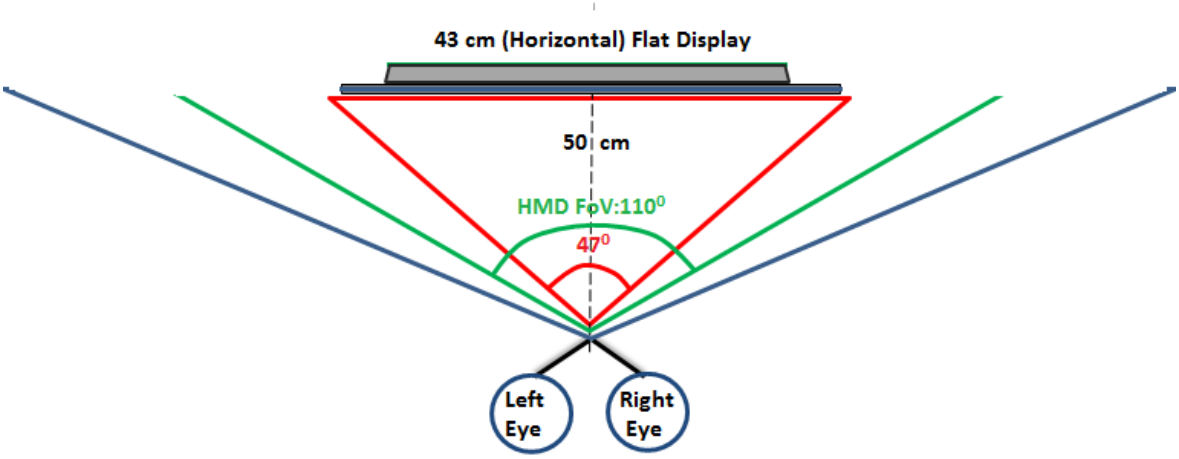


Figure 2.4. The schematics represents the approximate field of view for Oculus Rift Consumer Version compared to a standard medical display from 50 cm away.

2.3.4. Portability and Cost

In a standard radiology reading room, multiple conventional flat displays are used for the comparison aspect. The regular reading room with multiple medical grade monitors obliges an investment of space and cost. Virtual reality exceeds the limitation

of time and space. So, VR environments promote the collaboration in sharing space either physically or virtually. Telemedicine which is composed of telepresence,virtual reality and augmented reality gets great benefit from virtual reality techniques. This application enables physicians to diagnose and treat patients at a distance. Teleradiology is a convenient way of bringing well trained and experienced doctors into a virtual zone so that they can be accessible for even patients who may be located thousands of kilometers away.

3. LESION DETECTION

Every year, 1.69 million people are dying due to lung cancer in worldwide [14]. In the United Kingdom, approximately 35000 people which is the 6% of total national death are dying in a year because of lung cancer [15]. Lung cancer is responsible for the death of nearly 8 in 10000 people in Scotland [16]. Statistical mortality rates because of lung cancer are similar in other countries. In the United States, every year, around 160000 people are dying due to lung cancer [17]. Although lung cancer is a fatal disease, it can be treated if it is detected earlier. Stage 1, 85% of early-detected lung cancer case has a 5-year of survival rate. On the other hand, people that have advanced lung cancer die in less than 2 years [18]. When the lung cancer begins to threat the people's life, scientists try to find the ways of detecting the disease earlier in order to increase the remedy rate. Early detection of lung cancer mainly depends on imaging using the chest radiography and computed tomography (CT). In both imaging techniques, the principle depends on the absorption of X-ray beams passing through the human body. X-ray radiography provides 2D image with a small radiation dose (0.02-0.1 mSv) where as CT provides 3D image with higher radiation dose (about 5 mSv for one chest CT). Nodules usually show no symptoms. Since they are asymptomatic, they are generally discovered by a chance on a chest X-ray that is done for different reason. They are usually smaller than 3–4 cm in diameter. Most lung nodules are spherical solid shape tissue placed in the lung parenchyma, which is the portion of the lung involved in the gas exchange. In both imaging techniques, nodules appear as a "light dot" bounded by the normal dark lung tissue. However, all nodules do not mean that they are cancer. Some nodules are benign where as some are malignant. All lung nodules are mostly considered as a cancer among people. Whereas around 50% of surgically removed nodules end up with benign [19]. On the other hand, malignant nodules may be the lung cancer tumor itself or the secondary lung cancer, metastases. Metastasis is a cancer that comes from other parts of the body. The carcinoma of the pancreas, kidneys, thyroid gland, colon, stomach or breast mostly causes pulmonary metastases. Only 1% of small nodules that are smaller than 5 mm in diameter are malignant in patients without the cancer history. The risk of malignancy is directly related with the

size of the nodule [20]. On the other hand, approximately 50% of incidentally detected nodules that are greater than 8 mm in diameter are malignant [21]. Early detection and treatment of both primary and secondary lung cancer are essential to reduce the mortality rate. Consequently, it is significant to eliminate the radiologic problems of detection and the interpretation of radiographs. Some aspects such as observer error, lesion characteristics or technical considerations are responsible for missed lung cancer [22].

3.1. Causes of Missed Lung Cancer

3.1.1. Observer Error

The major cause of misinterpretations of lung carcinoma is observer error [23,24]. This can be altered by improving reader's ability [25].

According to Kundel [22], scanning error, recognition error, and decision-making error are the sub causes of observer error.

If the lung lesion is not focused on the fovea while analyzing image, scanning error arises. Approximately 30% of missed lung lesions are resulted in scanning error. Although humans have 180° field of view, when screening an image at a distance of 40 cm, only the circle sized 2.5 cm in diameter of the central part specifies vivid image details [26]. Additionally, the focusing period of fovea continues about 200–300 ms. After passing that period, eyes lose their fixations on target and move to a new place [26]. Misdiagnosing of lung lesion is inevitable because, most of the time, some parts of chest radiograph cannot be scanned with fovea by the radiologists.

Nearly 25% of misdiagnosis lesions are happened because of the recognition error. The major factor that causes this error is the radiologist's attitude [25]. Work condition, fatigue, presence of other abnormalities can affect the observer's approach to the case.

The decision-making error accounts for 45% of missed lesion. This error mostly depends on the radiologist's experience [27]. Radiologists may misinterpret the lesion as shown in Figure 3.1 [28].

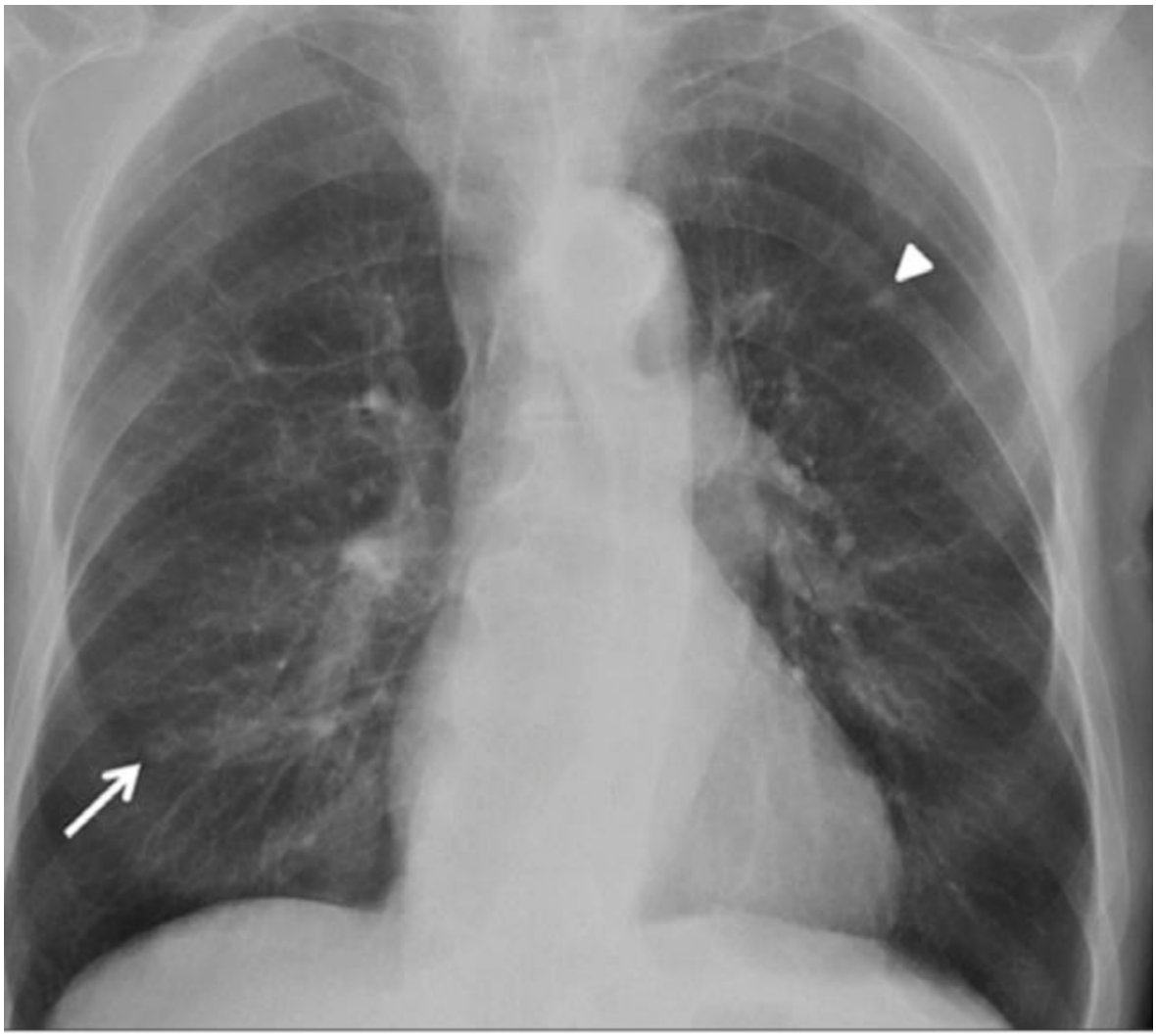


Figure 3.1. There is a nodular opacity on the left shown as an arrowhead. Nodular opacity shown as an arrow was not reported because it is considered as a nipple.

3.1.2. Lesion Characteristics

Lesion characteristics such as dimension, conspicuity, and location play important roles to get detected by the physicians. Dimension of the nodule is effective to recognize the tumor on chest X-ray. If the size of the nodule is getting bigger, the detection rate increases. Quekel confirms that nodules that are smaller than 10 mm have a detection error rate of 29%. If the lesions are 10-30 mm in size, the error rate is

28%. For the lesions from 30 to 40 mm, the rate is 12%, on the other hand if the lesions are greater than 40 mm, physicians do not miss them [29]. Conspicuity, which means the proportion of the contrast of the lesion and the structural complexity of the surrounding, is another critical feature to detect the lesion in the lung. Not only the visibility of the lesions' boundary affects the conspicuity but also the density of the tumor. If the tumors do not have sharp boundary, they are likely to be missed by the radiologists [29]. The location of the lesion is another criteria to contribute to the detectability. Austin [23] found out that 56% of missed pulmonary tumors are located mainly on the right lobes of the lung. Also, 81% of them are located in the upper lobes. There are many studies to support that the higher occurrence rate of the missed lung cancer are detected in the upper lobes [30]. Figure 3.2 shows an example of the missed lung cancer in a 75 years old man with pacemaker [28].

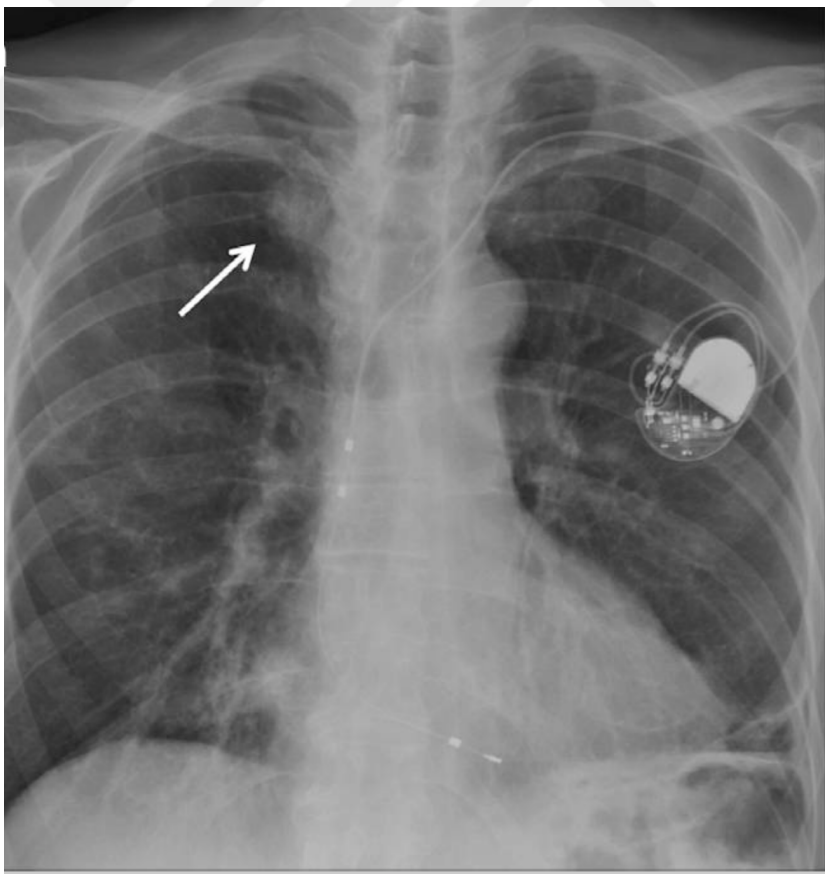


Figure 3.2. An oval opacity in the right upper lobe shown as an arrow was missed by the radiologist.

If the lesions are located in the blind spots of the lung such as apices, retrocardiac or hilar regions, they are more likely to be undetected by the readers because of the anatomic noise that is caused by the overlapping of anatomical structures of chest on each other on the radiographs.

3.1.3. Technical Concerns

Technical aspects such as image quality, patient positioning and movement also contribute to the detectability of the lung tumors. For the typical chest radiography, the front view (posteroanterior) and side view (latero-lateral) projections are obtained when the patient stands up. Some studies claim that lateral projection may be helpful to decrease the anatomical noise [31]. Austin *et al.* [23] finds out that a lateral radiograph shows the missed lesion better than the posteroanterior radiograph in four patients out of 23 patients (17%) . Wu *et al.* [32] reveals that lateral radiography is better than posteroanterior radiography in 8 out of 30 patients (27%) .

Instead, Forrest *et al.* [33] claims that lateral radiography is not needed. Every lung cancer is detectable on frontal X-ray.

3.2. Approaches to Decrease the Occurrence of Missed Pulmonary Malignancies on Chest Radiography

Reducing observer error and improving techniques can alter the prevalence of unseen pulmonary cancers.

3.2.1. Reducing Observer Error

Observer error mainly depends on the awareness and interpretation of lesions on radiographs. Ideal screening requires excellent screen, adjusting the distance of screen, and quiet work settings [34]. Teaching the chest reading methodology shown in Figure 3.3 also reduces the failure to recognize a lesion because of observer error [28, 35].

Comparison of previous X-ray radiograph and also side by side comparison of CT may help radiologists to understand the reasons behind the missed lung cancer issues [36].

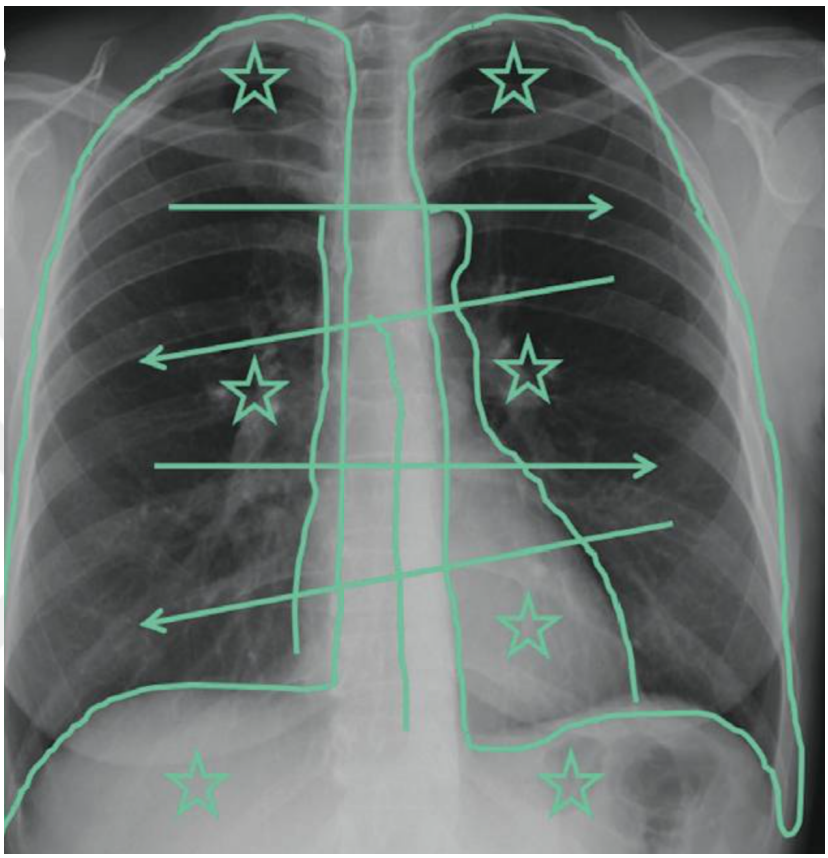


Figure 3.3. The representative figure of chest reading methodology. All lung zones (shown by arrows) are scanned. Stars show the "blind zones" and lines show the mediastinal lines and stripes.

3.2.2. Improving Tumor Detection Techniques

In the 1980s, the advancement in digital systems endorsed the growth of new tools to enhance the identification of unclear nodules. By using digitized X-ray, radiologists began to detect the lung lesions better. Image subtraction approaches may also increase lesion visibility. Image of two different tissues such as soft tissues and calcified structures are produced by dual-energy subtraction imaging with respect to the known energy dependence of X-ray attenuation of both tissues. After producing two images, the superimposing anatomic noise can be eliminated from the bones [37]. Another

approach contribute to the lesion visibility is the temporal subtraction imaging. In this technique, the previous radiograph of patient is subtracted from the current radiograph [38] . Thus, physicians have a chance to see the change in elusive malfunctions.

Recently, using computer aided detection system has been confirmed to increase the sensitivity of chest radiography. Some studies show that especially for the less experienced readers, using computer aided detection system creates more detection accuracy [39, 40]. White *et al.* [41] finds out that 50% of the nodules missed by the radiologists were detected by the computer aided detection system.

4. METHODOLOGY

4.1. ROC Analysis

The receiver operating characteristic (ROC) curve can be used as an effective method to assess the performance of diagnostic tests. ROC curve that is crucial for evaluating diagnostic ability of tests to compare two diagnostic tasks when each task is performed on the same subject [42]. It is extensively used in radiology for evaluating the performance of many radiological tests. The receiver operating characteristic (ROC) method was formerly used for radio signal detection and ranging (RADAR) in the 1950's. It was first applied in radiology in the 1960's [43]. ROC methods are specifically used to compare a new imaging technique with an existing 'gold-standard', when the actual disease status of the patient is identified [44].

The receiver operating characteristic (ROC) curve is defined as a plot of test sensitivity as the y coordinate versus its 1-specificity as the x coordinate [45]. The ROC curve can also be represented equivalently by plotting the fraction of true positives (TP = true positive rate) vs. the fraction of false positives (FP = false positive rate). The typical method of diagnostic test evaluation uses sensitivity and specificity as measures of accuracy of test [46].

The sensitivity shows whether the test is good or not according to the selected cases with the related symptom. It is also called the true positive rate. It can be defined as the ratio of number of true positive decisions and actually positive cases. Specificity is called the true negative rate, which is defined as the ratio of true negative decisions and the number of actually negative cases shown in Figure 4.1.

		Predicted Condition: Positive	Predicted Condition: Negative	Total
True Condition	Positive	True Positive (A)	False Negative (C)	A+C
	Negative	False Positive (B)	True Negative (D)	B+D
	Total	A+B	C+D	A+B+C+D

Figure 4.1. The classification table (counts).

True Positive Rate (TPR) or Sensitivity can be calculated by

$$A/(A+C) \tag{4.1}$$

and True Negative Rate (TNR) or Specificity can be calculated by

$$D/(B+D) \tag{4.2}$$

where A is True Positive, B is False Positive, C is False Negative and D is True Negative.

The false negative rate is the ratio of the cases with a known positive condition for which the predicted condition is negative. The false positive rate is the ratio of the cases with known negative cases for which the predicted condition is positive. This rate represents the X-axis on the ROC curve.

False Negative Rate (FNR) or Miss Rate is obtained by

$$C/(A+C) \tag{4.3}$$

False Positive Rate (FPR) or Fall-out is obtained by

$$B/(B+D) \tag{4.4}$$

where A is True Positive, B is False Positive, C is False Negative and D is True Negative.

If the test results indicate either the presence or absence of a symptom (one of two defined categories), then it means that there is only one pair of sensitivity and specificity values. But, in many diagnostic cases in medical imaging, image detection may not be noticeable. The confidence level variations among radiologists who interpret the findings during diagnostic evaluations should be taken into consideration. Consequently, a single pair of sensitivity and specificity values is not satisfactory to designate the performance of a diagnostic test [47].

A ROC curve can be evaluated by making the assumption of a binormal distribution (i.e., one of the two Gaussian distributions for the test results belongs to patients with benign nodules and the second one for the test results of those patients with malignant nodules) [48]. As a result, the fitted or smooth ROC curve determined by two parameters can be obtained as shown in Figure 4.2.

The first parameter is the standardized difference in the means of the distributions of the test results for those subjects with and without the symptom. The other parameter is the ratio of the standard deviations of the distributions of the test results for those subjects without versus those with the symptom [48]. ROC curve can be constructed also by connecting pair of values representing to the FPR and sensitivity

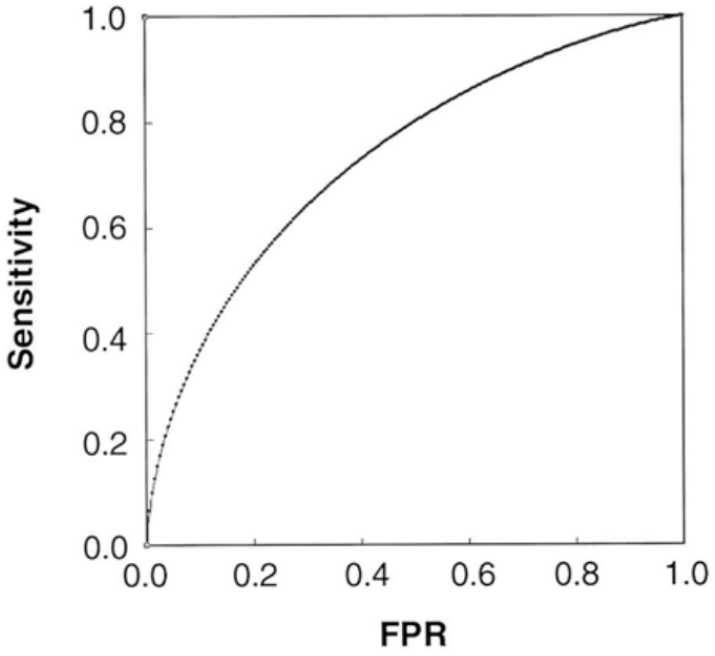


Figure 4.2. The fitted ROC curve (formed through the assumption of a binormal distribution).

acquired at all the possible cutoff levels between endpoints (0,0 and 1,1) respectively as shown in Figure 4.3 [49]. The resulting ROC curve is called the empirical ROC curve that demonstrates the relationship between sensitivity and FPR at all possible cutoff levels. Therefore, ROC curve can be used to evaluate the performance of a test without depending on the decision threshold.

The combined measurement of sensitivity and specificity is called Area Under Curve (AUC). AUC indicates the total performance of a diagnostic test. The average value of sensitivity for all possible values of specificity is the measure of AUC. AUC can range any value between 0 and 1. To be the better performance of a diagnostic test, the AUC must be closer to 1. A test that is measured to be absolutely perfect is shown in Figure 4.4 as a curve A that has an AUC value of 1. The chance discrimination that the AUC value is 0,5 is placed on diagonal line as curve D in Figure 4.4. The minimum AUC value of 0,5 is regarded as a chance level. A diagnostic test should have greater AUC value than 0,5 in order to be acceptable [49].

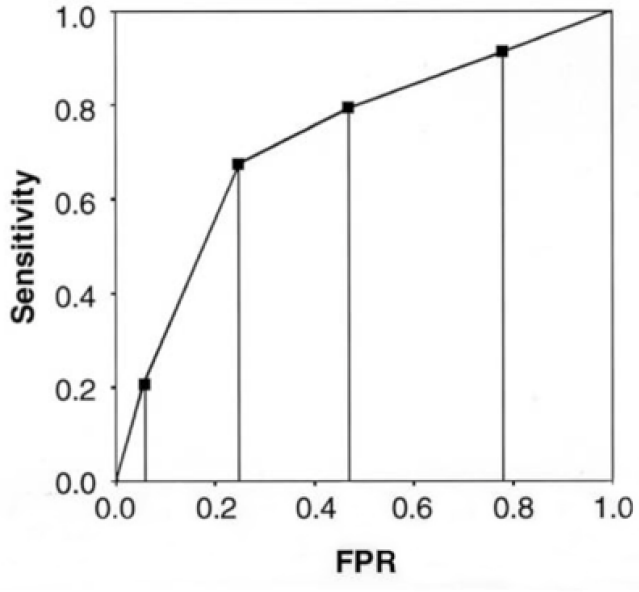


Figure 4.3. The empirical ROC curve. Dots on the graph represent the discrete points .

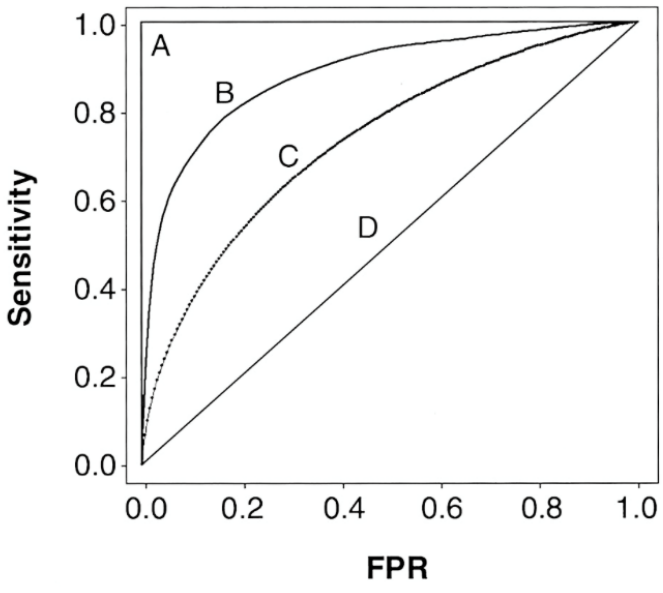


Figure 4.4. Four ROC curves. Curve A represents a perfect test (has an AUC of 1). ROC curves B and C can discriminate between those subjects with and without disease. Diagnostic performance of Test B (higher AUC) is better than test C. Curve D that is called as chance diagonal has an AUC of 0.5.

Sensitivity, specificity and AUC do not depend on disease prevalence [47]. AUC can be assessed both parametrically and nonparametrically by the binormal distribution and empirical ROC curve without any distributional assumption respectively [50].

The connecting points form the trapezoids .The total area of trapezoids represents the nonparametric evaluation of the area under the empirical ROC curve as shown in Figure 4.3 [49] .

However, in most ROC analysis of radiological tests, five or six categories (e.g., definitely no nodule, probably no nodule, uncertain, probably nodule and definitely nodule) discrete rating scales are used to prevent getting degenerate data for the parametric method.

AUC is often computed with 95% confidence interval. An AUC of a test obtained from a group of patients is not a fixed, true value, but a value from a sample that there can be a chance of statistical error. Thus, if different group with the same characteristics performs the same test, the obtained AUC may be different [51].

The true value of AUC is within the 95% confidence interval means there is a 5% chance of being wrong. Consequently, the test has statistically better performance if the lower bound of the 95% confidence interval of AUC for a test is greater than 0.5 [52].

One can match the performance of two different tests by comparing their AUCs. If the AUC is high then it has a better performance. On the other hand, even if the two tests generate the same overall diagnostic performance that means they have equal AUC values, the two ROC curves of the two tests do not mean they are identical [49]. Two ROC curves with equal AUCs shown in Figure 4.5 are not identical, despite the fact that the overall performances of the two tests are the same. Test A is better than test B in the low FPR range (or low sensitivity range) however test B is better than test A in the high FPR range (or high sensitivity range) shown in Figure 4.5.

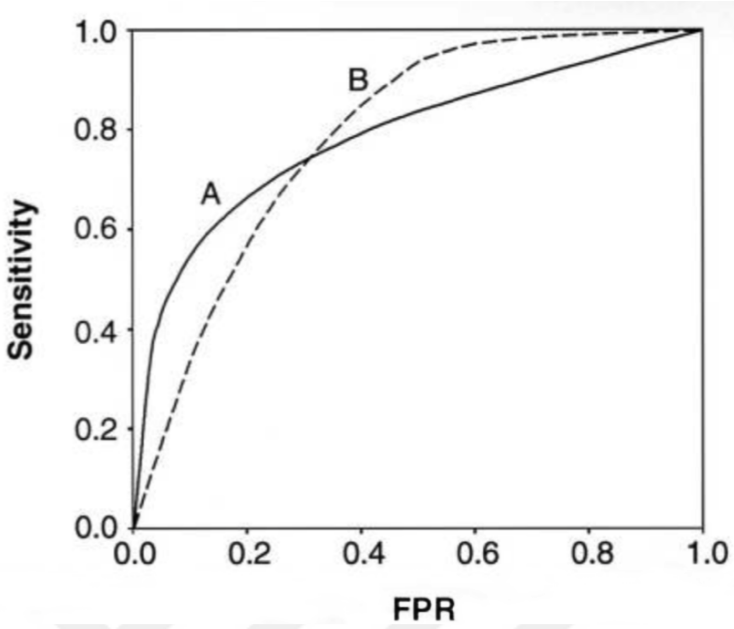


Figure 4.5. Although curve A and B have equal AUC, they are not identical. Test A has better sensitivity range however test B is better than test A in the high FPR range.

For comparing the performances of distinctive diagnostic tests, calculating only partial ROC area may make the diagnostic performance ambiguous. In screening for a severe disease such as cancer in a high-risk group, one can choose the cutoff range for a positive test to provide better sensitivity [47].

Instead, in screening for a specific disease that has a low occurrence rate, a high specificity and low FPR is vital. Changing the cutoff range for a positive test correspondingly can cause many needless, risky follow-ups and/or treatments for the patients. All positive decisions will be regarded as false positive decisions. In Figure 4.5 shows the total performances of the two tests are the same, but for diagnostic decision that requires high sensitivity, test B would be better.

It would be better to assess the performance of the diagnostic test according to the diagnostic situation. Considering only the portion area of the total ROC curve, the valuable pointers would be the partial ROC area and sensitivity at a particular FPR.

In this study, the test results were recorded in ordinal scale (e.g. 5 ordinal scale: "definitely no nodule", "probably no nodule", "uncertain", "probably nodule", "definitely nodule"), the sensitivity and specificity were computed for all the possible threshold values [51] by the web based calculator called as "jrocfit"(version 2.0) for ROC curves [53]. It contains JROCFIT and JLABROC4, JavaScript programs to calculate the receiver-operating characteristic (ROC) curves.

Through this calculator, the receiver operating characteristics curves for all radiologist - VR pairs and Medical monitor pairs and the statistical comparison of the data are stated in Chapter 5.

4.2. McNemar's Test

Quinn McNemar first introduced the McNemar test in 1947. The McNemar test can be considered as a type of chi-square test that uses dependent data rather than independent samples.

The McNemar's test is used for statistical significance of any difference in the image evaluation among the images sets within readers. If there are two paired variables that have exactly two possible outcomes as 0 and 1 each, the McNemar test can be used to test whether there is a statistically significant difference between the probability of a (0,1) pair and the probability of a (1,0) pair. This test is often used for the cases testing for the presence (= 1) or absence (= 0) of something and variable 1 (X_i) is the state before an experiment and variable 2 (Y_i) is the state after the experiment.

Table 4.1. Data summary of 2x2 matrix of McNemar Test.

	Y_i=**0**	Y_i=**1**
X_i=**0**	a	b
X_i=**1**	c	d

McNemar test assumes that the pairs (X_i,Y_i) are mutually independent. Each X_i and Y_i can be assigned to one of two possible categories. We compared the two-tailed p value of the VR system with standard DICOM viewer and medical monitor by calculating them through McNemar test. McNemar test was done by using XLSTAT-Base software [54].

4.3. Kappa Testing

Kappa test is a nonparametric test that can be used to measure inter/intra-observer agreement on imaging studies. The interpretations that vary between and among two or more readers are called inter-observer whereas the interpretations completed by the same reader at separate time points are called intra-observer. Reader agreement studies [55] are needed to confirm internal validity in studies where there is more than one reader. In comparing variability between continuous measurements, Cohen kappa statistic, which is a measure of reader agreement responsible for chance alone, is used for categoric variables [56].

When more than two categories of classification are possible, a weighted kappa statistic can be used [56].

When the outcomes are ordinal, ranked, subject to outliers or measured roughly, nonparametric tests are considered to be relatively simple to conduct.

The intra-agreement weighted kappa values were calculated by MedCalc software [57]. The summary of Kappa values and the level of agreement are given in the Table 4.2 [58].

Table 4.2. The summary of Kappa values and the level of agreement [58].

.

<=0	no agreement
0.01 - 0.20	none to slight
0.21- 0.40	fair
0.41 - 0.60	moderate
0.61 - 0.80	substantial
0.81 - 1.00	almost perfect agreement

5. PROTOTYPE DESIGN

5.1. Display

The Oculus Rift consumer version 2016 was chosen as the virtual reality head-mounted display to be used for the study. Consumer version was released by Oculus VR [10] in 2016.

Interpupillary distance (IPD) can be adjusted between 58 and 71 mm via the dial on the bottom of head mounted display. Foam faceplates are removable and interchangeable. This feature makes Oculus more comfortable for the users with glasses. Recommended frame width is 142 mm and a frame height of 50 mm for the users that have to wear glasses.

The resolution of Oculus Rift is 2160x1200 pixels (1080x1200 per eye). Organic light emitting diode (OLED) panel enlarge the image that provides a stereoscopic 3D view for users. The field of view is nearly 110° which is greater than the field of view acquired from a standard smartphone or tablet. The dual OLED panel displays of consumer version 1 (CV1) makes the color more vibrant. The resolution of the CV1 is not satisfactory enough to completely elude "screen door effect" which can be defined as the ability of distinguishing the individual pixels of the display. Another weakness of CV1's display and optics system is the "god rays" caused by Fresnel lenses' edges. The light scattered from the edge of lenses seems to be lens flares and are noticeable on the dark background. The CV1 has 2ms frame panel pixels with a high refresh rate of 90 Hz. The usage of a low persistence display with high refresh rate has some advantages such as significantly reduced motion blur and potentially less nausea and smaller details to be viewable by the user. This readability feature may enable a user to evaluate elusive changes between multiple images.

The tracking system is optical based called constellation. The Consumer Version 1 comes with 6 degrees of freedom (both rotational and positional) for head tracking and

sub-millimeter accuracy. While rotational tracking is achieved with inertial measuring unit (IMU) that is a combination of accelerometer, gyroscope, and a magnetometer to correct for float, an Oculus Sensor is responsible for positional tracking. The camera sensor is able to track LED markers located on the back straps of the HMD viewable in Figure 5.1.



Figure 5.1. Oculus Rift Consumer Version infrared markers.

The basic technical comparison of the Medical Monitor (Barco MDRC-2124) and Oculus Rift CV1 is specified in Table 5.1.

5.2. Software

The Medical Image Processing, Analysis, and Visualization (MIPAV) application is a free and open-source software. The overall package has been used for a variety of medical applications so far. Medical images of numerous modalities such as Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Computer Tomog-

Table 5.1. The technical comparison of medical monitor and Oculus Rift CV1.

Monitor Type	**Medical Monitor**	**Oculus Rift CV1**
Color Tone	Color	Color
Resolution	1920 x 1200 Pixels	2160 x 1200 Pixels
Weight	15.4 kg	470 g
DICOM Viewer	Radiant	MicroDicom
Price	1899 USD	599 USD

raphy (CT), or microscopy can be analyzed by MIPAV quantitatively. MIPAV is used to visualize the medical Digital Imaging and Communication in Medicine (DICOM) images and medical images were saved as DICOM images. Also, virtual desktop application is used to make the computer desktop to be seen through Oculus Rift CV1. Virtual desktop application enables to resize the computer's window in VR and also arrange the distance between the images and the users.

5.3. System Design

The system design is shown in Figure 5.2. Lung X-ray images were downloaded from the database to a high-end Oculus Rift/VR Ready laptop computer (MSI GT80 Titan SLI). DICOM images were reviewed using MIPAV software. 50 images with nodule, 50 images with non-nodule in total 100 X-ray images were chosen according to their difficulty level of being detected. Four datasets each includes 25 images randomly were prepared. Each image was numbered from 1-100 and saved as DICOM files separately.

Virtual desktop application was used to enlarge the computer window that is viewable via Oculus Rift CV1. By the help of this application, radiologists were able to arrange both the screen size and the distance between the window and themselves.

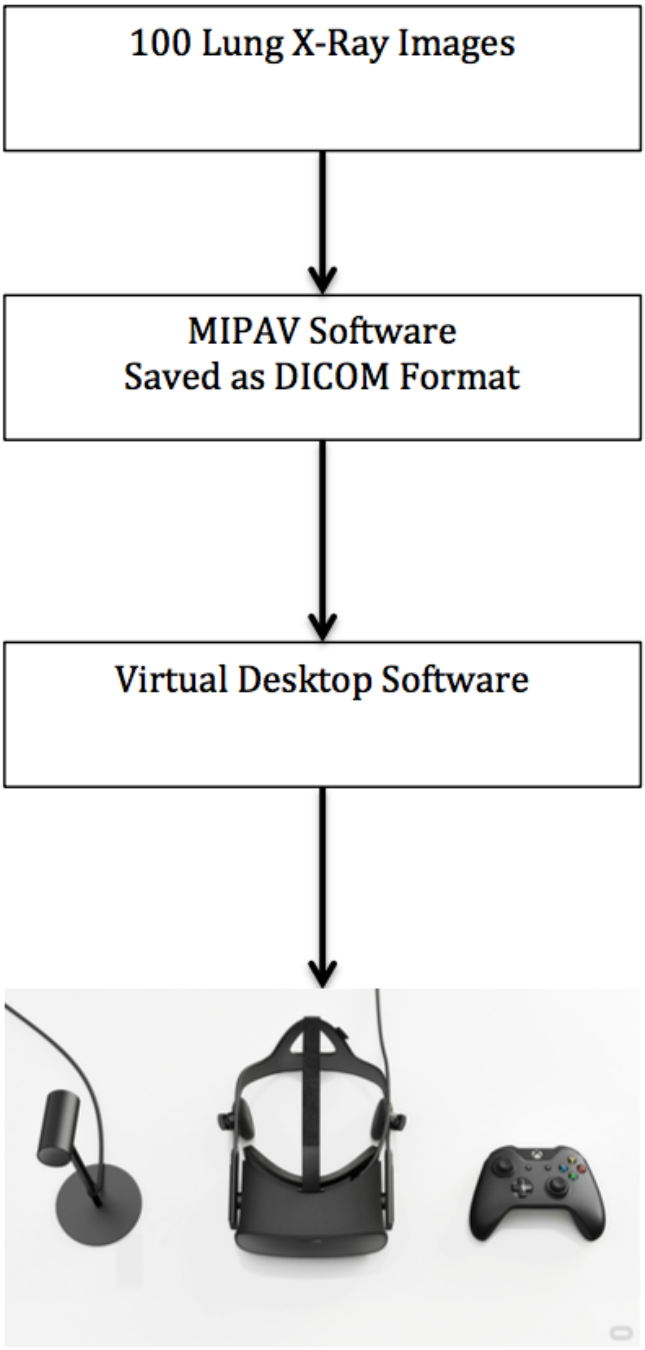


Figure 5.2. The system design layout with Oculus Rift HMD 2016.

5.4. Assessment of Display Performance

Geometric distortions begin from deviations that cause the displayed image to be geometrically different from the original image [59].

Such distortions can cause some consequences such as affecting the shapes of image characteristics and its relative sizes, mostly for larger displays.

Spatial accuracy for geometric distortion of a VR display system was visually determined using the TG18-QC (Figure 5.3) [60]. In this image, the small squares with dashed lines define the four quadrants of the pattern, there is a larger square at the center surrounding the luminance patches and it is for characterization of geometric distortions within the overall image.

The lines were straight that is an indicator of accurate linearity, without any curvature or unevenness.

In using the patterns, patterns in DICOM and 16-bit TIFF formats were displayed with a window and level that is set to cover the range from 0 to 4095 (WW = 4096, WL = 2048).

Display resolution was evaluated by visually considering the appearance of the CX patterns in the TG18-CX test patterns as seen in Figure 5.4 that is recommended by the American Association of Physicist in Medicine (AAPM) [60].

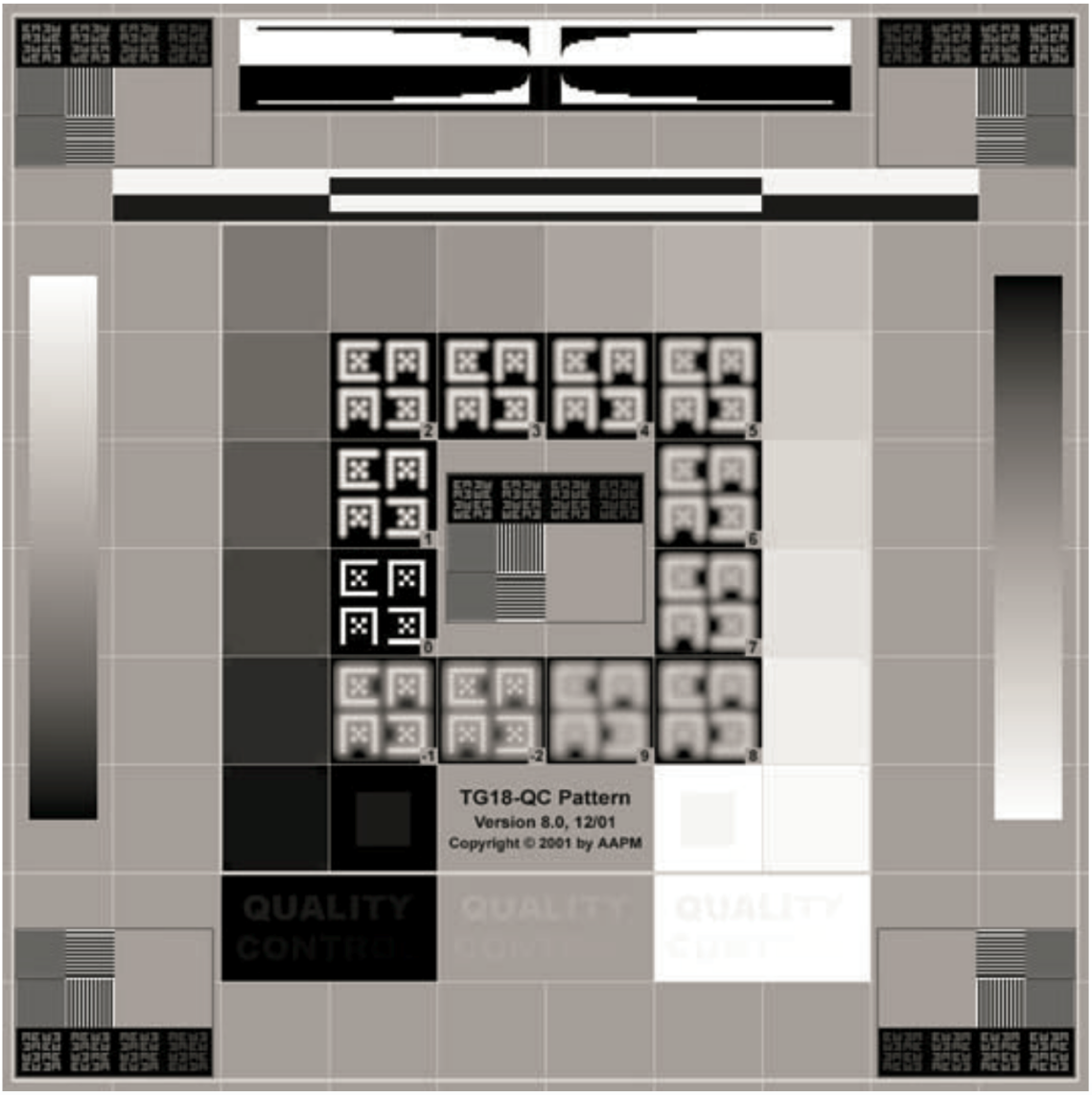


Figure 5.3. TG18-QC test pattern. It is used for quantitative assessment of geometric distortions.

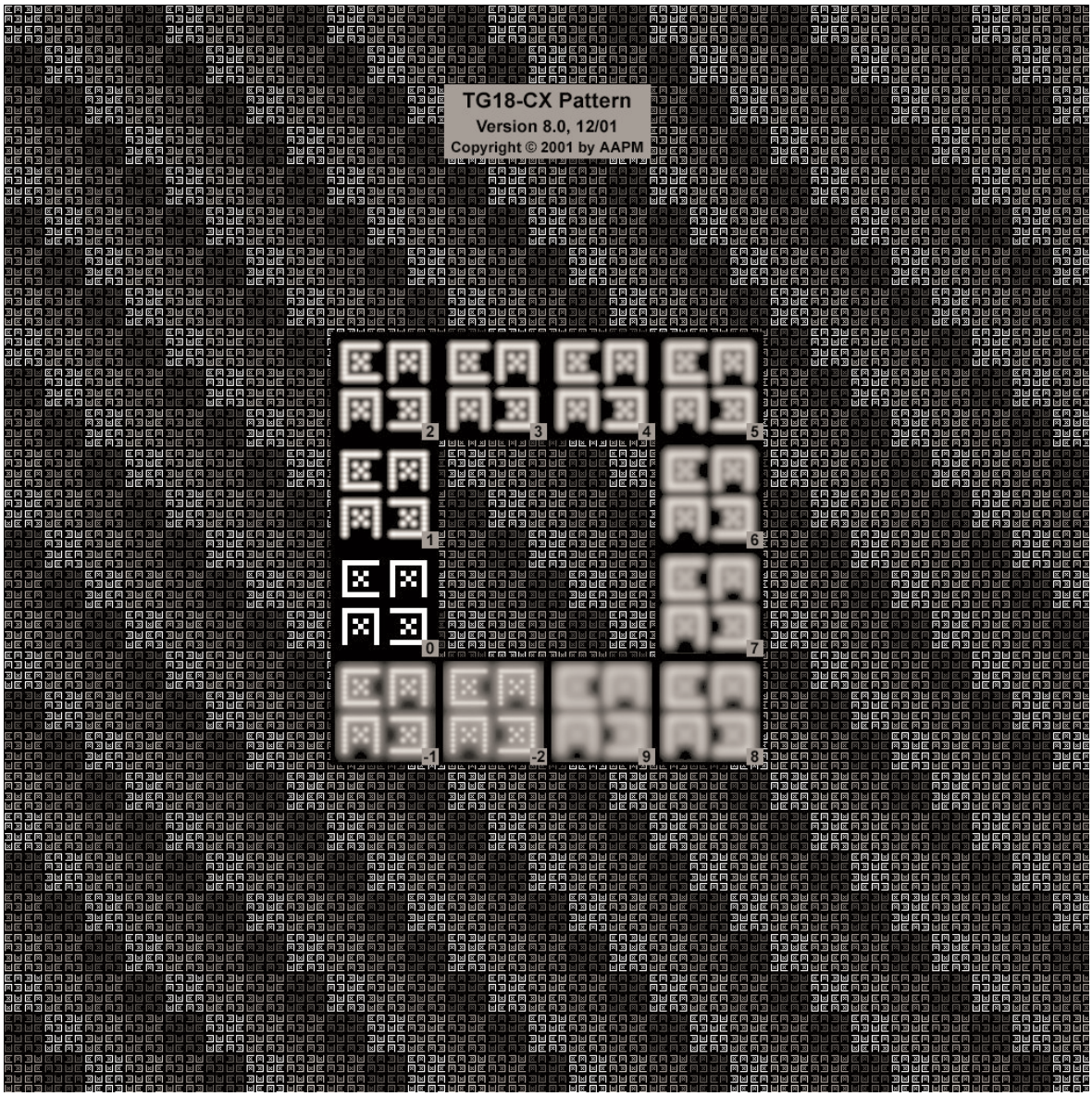


Figure 5.4. The TG18-CX test pattern for the assessment of display resolution and resolution uniformity.

6. EXPERIMENTS AND RESULTS

6.1. Differential Diagnosis Experiment

In this section, the design and results of an experiment evaluating usage of the virtual reality diagnostic imaging system for differential diagnosis, particularly for X-ray lung nodule identification and assessment is discussed.

6.1.1. Experiment Design

The experiments have been conducted in between March and May 2017 with the purpose of determining the effectiveness of the virtual reality diagnostic imaging prototype in allowing radiologists to both identify nodule and locate the lung nodules.

Five radiologists from Yeditepe University Radiology Department were enrolled in this study. They examined chosen 100 non-contrast-enhanced lung X-ray images downloaded from online database [61] by using the Oculus Rift 2016 VR head mounted display and also a medical monitor (MM) of Digital Imaging and Communication in Medicine (DICOM) viewer, scoring the probability of nodule on each exam on a five-point scale. The reviewers both graded the possibility of nodule in each case using a five-point scale (1 = definitely no nodule, 2 = probably no nodule, 3 = uncertain , 4 = probably nodule, 5 = definitely nodule) and stated the location of the nodule. After downloading images from the database, these images were saved in DICOM format by using Medical Image Processing, Analysis, and Visualization (MIPAV) software [62].

Lung X-ray images prepared to be seen from the Oculus Rift Virtual Reality Head Mount Display (VR HMD) by virtual desktop software program [63].

For each radiologist, 4 datasets that have 25 X-Ray lung images each were randomly assigned in order to be seen using the virtual reality diagnostic imaging system with MicroDicom [64], a free-to-use DICOM viewer. We randomly assigned each case a

number from 1 to 100. Fifty of X-ray images have elusive radiological signs of nodule; where as the other 50 show no significant abnormalities. There was no time limit. However, mandatory resting period of 10 minutes was insisted after reviewing 25 cases and another 30 minutes mandatory resting period was applied after 50 cases.

The experiment tasks and the instruction of Virtual Reality head mounted display system operation were explained to the participants that were assigned to use the virtual reality. Radiologists were able to use all functions of the DICOM viewer on VR head mounted display, including zooming, panning, window leveling and control of window width easily.

Radiologists then wore the Oculus Rift consumer edition virtual reality head-mounted display as shown in Figure 6.1.



Figure 6.1. Evaluation of lung X-ray images through Oculus Rift VR HMD.

The participants first asked to evaluate a trial image using the VR system to get use to the headset until achieving an optimal focus and clarity. When the participants got familiar with the VR system, a random dataset was displayed. The participants then assessed lung X-ray images in accordance with the nodule in the dataset and if any located a nodule as left upper, left lower, right upper and right lower. Participants

continued to select the other images in the dataset locating the nodule in each if there is any.

6.1.2. Experiment Results

In total, 5 participants assessed 100 image datasets (50 with nodules, 50 with none-nodule) with the VR system and a standard DICOM viewer. After one-month washout period, participants evaluated the same 100 lung X-ray images with the Medical monitors as in Figure 6.2.

The total evaluation time of only one reader for 25 lung X-ray images both through VR and medical monitor from the third data set was recorded. The third data set was chosen so that the radiologists could have enough time to get familiar with VR. Time consumed by using VR and medical monitor (MM) was recorded as 995 seconds (16.58 min) and 855 seconds (14.25 min) respectively in Table 6.1.

Table 6.1. Time consumed while assessing 25 X-ray lung nodules on VR and medical monitor (MM).

Consumed Time	**VR**	**MM**	**Difference**
in seconds	995	855	140
in minutes	16.58	14.25	2.33

Reader 1 evaluation results using both Oculus Rift VR and medical monitor were given in Figure 6.3.

Table 6.2 shows the specificity and sensitivity of datasets assessed by Reader 1 for each display.

Figure 6.4 shows the ROC curves of VR and medical monitor for Reader 1.

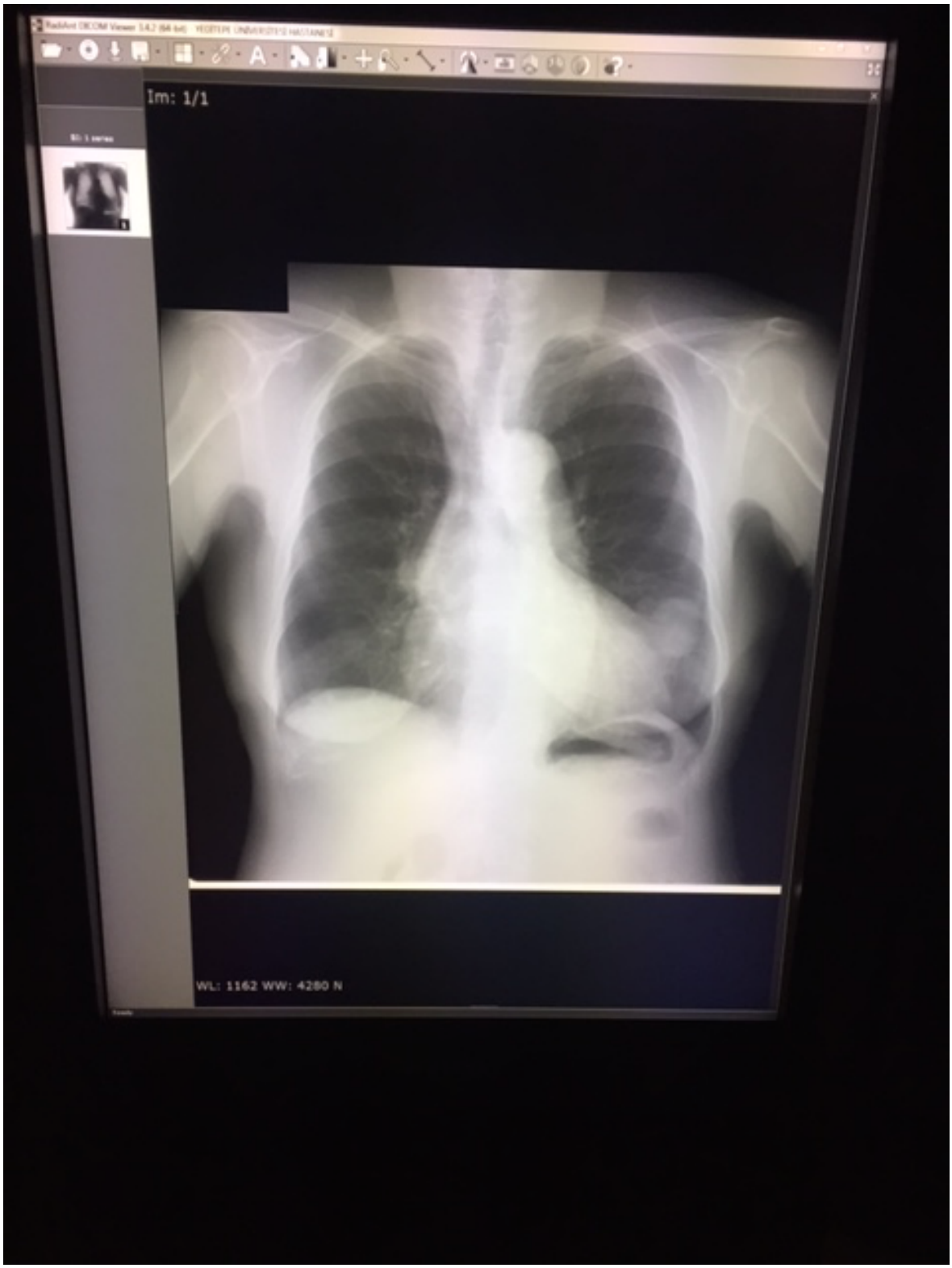


Figure 6.2. A sample X-ray lung image from the database seen through medical monitor.

Case Number	Nodule Existence in reality	Reader 1 VR rating	Reader 1 MM Rating	Case Number	Nodule Existence in reality	Reader 1 VR rating	Reader 1 MM Rating
1	1	5	5	51	0	1	1
2	1	5	5	52	1	1	5
3	0	1	1	53	0	1	1
4	1	3	4	54	0	1	1
5	0	5	2	55	1	4	5
6	0	1	1	56	0	1	1
7	1	4	5	57	1	5	3
8	1	5	5	58	1	5	5
9	1	1	3	59	0	1	1
10	0	1	1	60	0	2	3
11	1	1	2	61	0	5	4
12	1	5	2	62	1	5	5
13	1	5	4	63	0	1	1
14	0	1	1	64	1	5	5
15	0	1	1	65	0	1	1
16	0	1	1	66	0	1	2
17	1	2	4	67	1	5	5
18	1	3	2	68	0	1	1
19	1	3	2	69	1	5	5
20	1	5	5	70	1	5	3
21	1	5	5	71	1	5	5
22	0	1	1	72	0	1	1
23	1	5	4	73	0	1	3
24	1	5	4	74	1	5	5
25	0	1	1	75	0	1	1
26	1	5	5	76	0	1	1
27	1	1	3	77	0	1	1
28	1	5	5	78	1	1	2
29	0	1	2	79	0	1	2
30	0	1	1	80	0	3	2
31	0	1	1	81	0	1	1
32	1	2	4	82	1	5	4
33	1	5	5	83	0	1	1
34	1	1	2	84	0	1	1
35	0	1	1	85	0	3	4
36	1	5	5	86	1	5	5
37	0	1	1	87	0	1	3
38	1	2	3	88	0	1	4
39	1	5	5	89	0	5	2
40	0	1	2	90	0	1	1
41	0	1	1	91	1	5	5
42	1	5	5	92	0	1	1
43	1	1	3	93	0	1	1
44	1	5	5	94	1	5	5
45	1	5	4	95	0	1	3
46	0	1	1	96	1	5	4
47	1	1	2	97	0	1	1
48	0	1	2	98	1	3	2
49	1	3	2	99	1	1	3
50	0	1	1	100	0	1	1

Figure 6.3. Reader 1's evaluation of lung X-Ray Images on VR and MM.

Table 6.2. Statistical comparison of Reader 1's evaluation on VR and MM.

	Reader 1-VR	**Reader 1- MM**
Number of Cases	100	100
Number of Correct	83	84
Accuracy	83%	84%
Sensitivity	76%	82%
Specificity	90%	86%
Positive Cases Missed	12	9
Negative Cases Missed	5	7
Fitted ROC Area	0.931	0.948
Empiric ROC Area	0.864	0.941
Std. Dev. (Area)	0.029	0.022

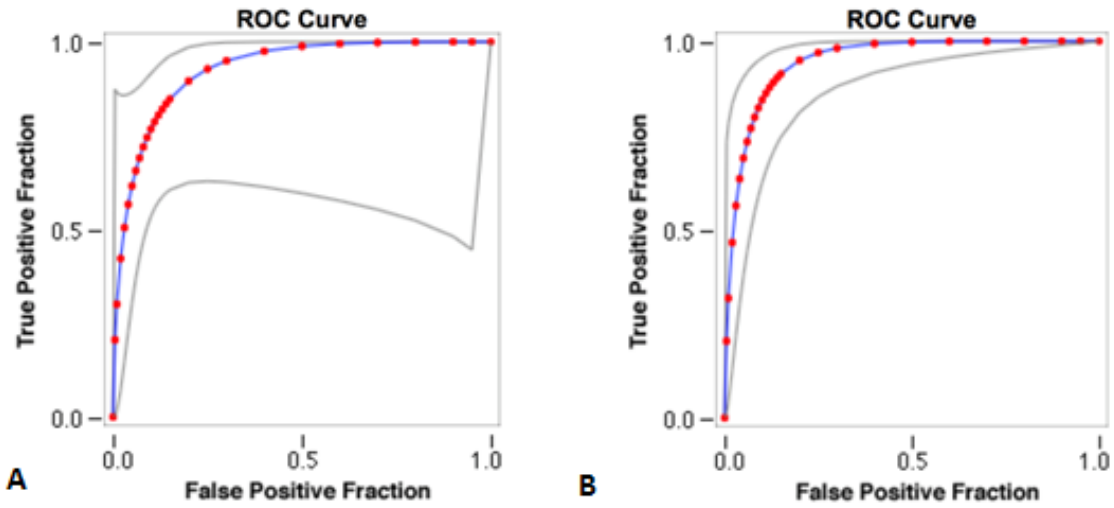


Figure 6.4. A: ROC curve of VR and B: ROC curve of MM for the Reader 1.

Reader 2 evaluation results using both Oculus Rift VR and medical monitor were given in Figure 6.5.

Table 6.3 shows the specificity and sensitivity of datasets assessed by Reader 2 for each display.

Figure 6.6 shows the ROC curves of VR and medical monitor for Reader 2.

Table 6.3. Statistical comparison of Reader 2's evaluation on VR and MM.

	Reader 2-VR	**Reader 2- MM**
Number of Cases	100	100
Number of Correct	78	84
Accuracy	78%	84%
Sensitivity	76%	82%
Specificity	80%	86%
Positive Cases Missed	12	9
Negative Cases Missed	10	7
Fitted ROC Area	0.865	0.915
Empiric ROC Area	0.797	0.882
Std. Dev. (Area)	0.046	0.037

Reader 3 evaluation results using both Oculus Rift VR and medical monitor were given in Figure 6.7.

Table 6.4 shows the specificity and sensitivity of datasets assessed by Reader 3 for each display.

Figure 6.8 shows the ROC curves of VR and medical monitor for Reader 3.

Case Number	Nodule Existence in reality	Reader 2 VR rating	Reader 2 MM rating	Case Number	Nodule Existence in reality	Reader 2 VR rating	Reader 2 MM rating
1	1	5	5	51	0	1	1
2	1	5	5	52	1	1	1
3	0	1	1	53	0	1	1
4	1	1	1	54	0	3	2
5	0	5	2	55	1	5	5
6	0	1	1	56	0	5	5
7	1	5	5	57	1	5	5
8	1	5	5	58	1	5	5
9	1	4	5	59	0	1	1
10	0	1	1	60	0	1	1
11	1	1	4	61	0	1	1
12	1	5	5	62	1	5	5
13	1	1	4	63	0	1	1
14	0	1	1	64	1	5	5
15	0	1	1	65	0	1	1
16	0	5	3	66	0	1	1
17	1	4	5	67	1	5	5
18	1	1	2	68	0	1	1
19	1	5	5	69	1	5	5
20	1	5	5	70	1	5	5
21	1	5	5	71	1	5	5
22	0	1	1	72	0	1	1
23	1	5	5	73	0	1	1
24	1	1	1	74	1	5	2
25	0	1	1	75	0	5	1
26	1	5	5	76	0	5	2
27	1	5	5	77	0	1	3
28	1	5	5	78	1	1	1
29	0	1	1	79	0	1	3
30	0	1	1	80	0	3	1
31	0	1	1	81	0	1	1
32	1	4	5	82	1	5	5
33	1	5	5	83	0	1	1
34	1	1	1	84	0	1	1
35	0	1	1	85	0	1	1
36	1	5	5	86	1	5	5
37	0	1	1	87	0	1	3
38	1	4	5	88	0	1	1
39	1	5	5	89	0	1	1
40	0	3	1	90	0	1	1
41	0	1	1	91	1	5	5
42	1	5	5	92	0	1	1
43	1	1	1	93	0	1	1
44	1	5	5	94	1	1	1
45	1	5	5	95	0	5	5
46	0	1	1	96	1	5	5
47	1	5	5	97	0	5	5
48	0	1	1	98	1	1	4
49	1	1	4	99	1	5	5
50	0	1	1	100	0	1	1

Figure 6.5. Reader 2's evaluation of lung X-Ray images on VR and MM.

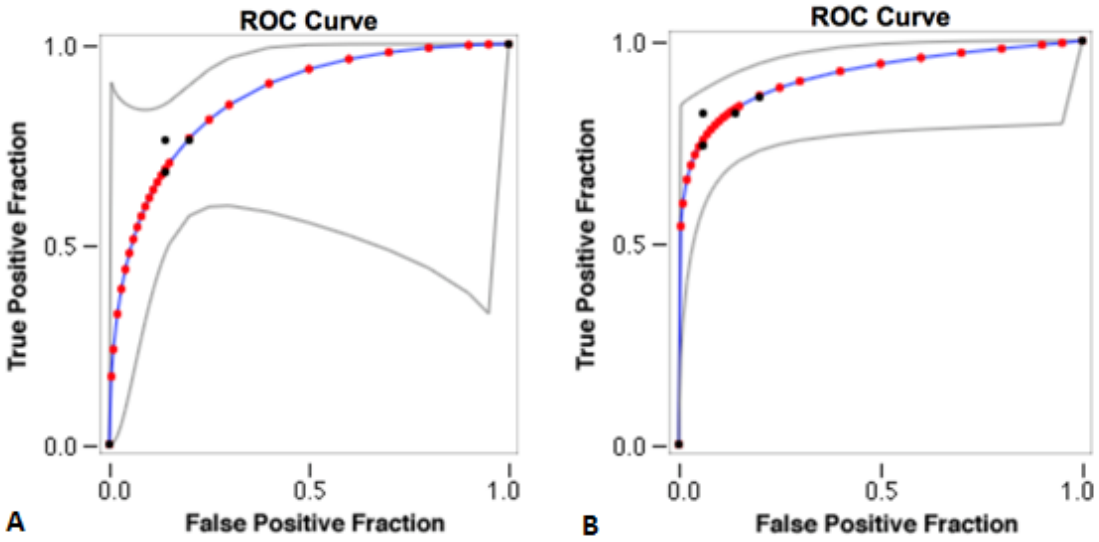


Figure 6.6. A: ROC curve of VR and B: ROC curve of MM for the Reader 2.

Table 6.4. Statistical comparison of Reader 3's evaluation on VR and MM.

	Reader 3-VR	**Reader 3- MM**
Number of Cases	100	100
Number of Correct	80	77
Accuracy	80%	77%
Sensitivity	72%	76%
Specificity	88%	78%
Positive Cases Missed	14	12
Negative Cases Missed	6	11
Fitted ROC Area	0.805	0.897
Empiric ROC Area	0.797	0.884
Std. Dev. (Area)	0.046	0.033

Case Number	Nodule Existence in reality	Reader 3 VR rating	Reader 3 MM rating	Case Number	Nodule Existence in reality	Reader 3 VR rating	Reader 3 MM rating
1	1	5	5	51	0	1	1
2	1	5	5	52	1	1	2
3	0	2	1	53	0	1	1
4	1	4	5	54	0	2	3
5	0	4	2	55	1	2	3
6	0	2	4	56	0	2	3
7	1	5	5	57	1	5	5
8	1	2	4	58	1	5	5
9	1	1	4	59	0	2	1
10	0	2	3	60	0	1	2
11	1	2	3	61	0	2	1
12	1	5	5	62	1	5	5
13	1	5	5	63	0	2	1
14	0	1	1	64	1	4	2
15	0	1	1	65	0	1	1
16	0	1	1	66	0	1	1
17	1	2	3	67	1	4	5
18	1	3	4	68	0	2	3
19	1	3	4	69	1	5	5
20	1	1	1	70	1	5	5
21	1	4	2	71	1	5	5
22	0	1	3	72	0	1	2
23	1	4	5	73	0	2	2
24	1	3	4	74	1	3	4
25	0	5	4	75	0	2	2
26	1	5	5	76	0	2	1
27	1	2	2	77	0	2	1
28	1	5	2	78	1	1	2
29	0	1	1	79	0	2	1
30	0	2	2	80	0	4	3
31	0	2	2	81	0	4	2
32	1	5	5	82	1	4	3
33	1	5	5	83	0	2	3
34	1	1	3	84	0	2	3
35	0	1	1	85	0	2	1
36	1	5	5	86	1	4	5
37	0	1	1	87	0	2	1
38	1	1	1	88	0	2	1
39	1	2	3	89	0	1	2
40	0	1	1	90	0	2	1
41	0	2	1	91	1	4	5
42	1	5	5	92	0	4	3
43	1	2	2	93	0	1	1
44	1	5	5	94	1	4	5
45	1	5	5	95	0	5	2
46	0	1	2	96	1	3	2
47	1	5	4	97	0	1	2
48	0	2	1	98	1	4	5
49	1	4	2	99	1	1	2
50	0	1	1	100	0	2	1

Figure 6.7. Reader 3's evaluation of lung X-Ray images on VR and MM.

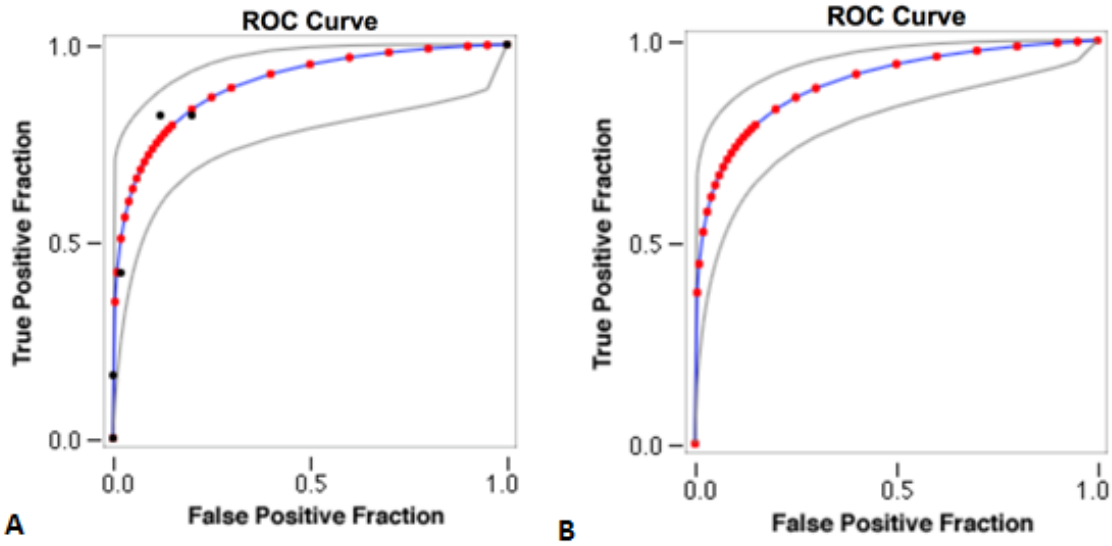


Figure 6.8. A: ROC curve of VR and B: ROC curve of MM for the Reader 3.

Table 6.5. Statistical comparison of Reader 4's evaluation on VR and MM.

	Reader 4-VR	**Reader 4- MM**
Number of Cases	100	100
Number of Correct	72	75
Accuracy	72%	75%
Sensitivity	76%	74%
Specificity	68%	76%
Positive Cases Missed	12	13
Negative Cases Missed	16	12
Fitted ROC Area	0.775	0.771
Empiric ROC Area	0.723	0.747
Std. Dev. (Area)	0.055	0.052

Case Number	Nodule Existence in reality	Reader 4 VR rating	Reader 4 MM rating	Case Number	Nodule Existence in reality	Reader 4 VR rating	Reader 4 MM rating
1	1	5	4	51	0	1	1
2	1	5	5	52	1	1	2
3	0	1	5	53	0	5	4
4	1	2	1	54	0	1	4
5	0	3	2	55	1	5	4
6	0	2	2	56	0	1	2
7	1	5	4	57	1	5	4
8	1	5	4	58	1	5	5
9	1	1	4	59	0	1	2
10	0	3	2	60	0	5	2
11	1	1	2	61	0	1	2
12	1	5	4	62	1	5	4
13	1	5	4	63	0	1	2
14	0	1	2	64	1	5	4
15	0	1	1	65	0	1	2
16	0	1	2	66	0	5	2
17	1	4	4	67	1	5	4
18	1	5	4	68	0	2	2
19	1	5	5	69	1	5	5
20	1	5	4	70	1	5	5
21	1	5	4	71	1	5	5
22	0	1	2	72	0	1	2
23	1	5	2	73	0	1	2
24	1	3	2	74	1	1	4
25	0	5	5	75	0	5	4
26	1	5	5	76	0	5	2
27	1	1	4	77	0	5	2
28	1	5	4	78	1	1	2
29	0	1	2	79	0	5	4
30	0	2	2	80	0	1	2
31	0	5	2	81	0	1	4
32	1	5	2	82	1	5	5
33	1	4	4	83	0	1	4
34	1	1	2	84	0	1	2
35	0	5	5	85	0	1	2
36	1	5	5	86	1	5	4
37	0	1	4	87	0	1	2
38	1	1	2	88	0	1	2
39	1	1	2	89	0	5	2
40	0	5	5	90	0	1	2
41	0	5	2	91	1	5	4
42	1	5	5	92	0	1	4
43	1	1	2	93	0	1	2
44	1	5	4	94	1	5	4
45	1	5	5	95	0	1	2
46	0	1	2	96	1	5	2
47	1	1	2	97	0	5	2
48	0	1	2	98	1	5	5
49	1	5	4	99	1	4	4
50	0	1	2	100	0	1	2

Figure 6.9. Reader 4's evaluation of lung X-Ray images on VR and MM.

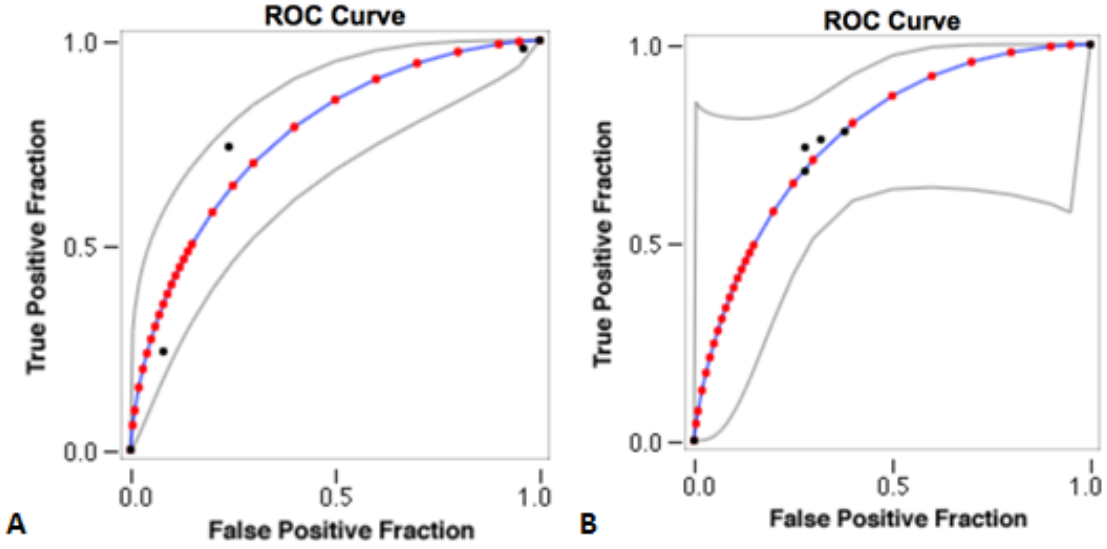


Figure 6.10. A: ROC curve of VR and B: ROC curve of MM for the Reader 4.

Reader 4 evaluation results using both Oculus Rift VR and medical monitor were given in Figure 6.9.

Table 6.5 shows the specificity and sensitivity of datasets assessed by Reader 4 for each display.

Figure 6.10 shows the ROC curves of VR and medical monitor for Reader 4.

Reader 5 evaluation results using both Oculus Rift VR and medical monitor were given in Figure 6.11.

Table 6.6 shows the specificity and sensitivity of datasets assessed by Reader 5 for each display.

Figure 6.12 shows the ROC curves of VR and medical monitor for Reader 5.

Case Number	Nodule Existence in reality	Reader 5 VR rating	Reader 5 MM rating	Case Number	Nodule Existence in reality	Reader 5 VR rating	Reader 5 MM rating
1	1	4	4	51	0	1	4
2	1	4	5	52	1	1	1
3	0	1	1	53	0	1	3
4	1	3	1	54	0	3	1
5	0	3	1	55	1	1	3
6	0	1	4	56	0	3	1
7	1	1	1	57	1	3	4
8	1	5	1	58	1	4	5
9	1	3	1	59	0	3	1
10	0	3	1	60	0	1	1
11	1	1	1	61	0	4	3
12	1	5	3	62	1	4	5
13	1	3	3	63	0	1	1
14	0	4	1	64	1	3	3
15	0	1	1	65	0	1	1
16	0	1	3	66	0	1	1
17	1	4	3	67	1	3	1
18	1	3	3	68	0	1	1
19	1	3	4	69	1	5	5
20	1	3	1	70	1	4	5
21	1	1	5	71	1	4	5
22	0	1	3	72	0	1	1
23	1	3	1	73	0	1	1
24	1	1	1	74	1	1	3
25	0	1	5	75	0	2	1
26	1	5	5	76	0	1	1
27	1	3	3	77	0	3	4
28	1	5	5	78	1	1	1
29	0	1	1	79	0	3	1
30	0	1	2	80	0	1	1
31	0	3	1	81	0	1	1
32	1	1	3	82	1	3	5
33	1	4	5	83	0	1	1
34	1	1	3	84	0	1	3
35	0	1	5	85	0	1	1
36	1	3	4	86	1	3	4
37	0	3	1	87	0	1	1
38	1	1	5	88	0	1	1
39	1	1	3	89	0	1	1
40	0	1	1	90	0	1	1
41	0	3	3	91	1	1	1
42	1	4	5	92	0	3	4
43	1	4	3	93	0	4	1
44	1	3	4	94	1	1	4
45	1	3	4	95	0	1	1
46	0	4	1	96	1	3	1
47	1	1	1	97	0	1	1
48	0	1	4	98	1	1	4
49	1	3	1	99	1	3	3
50	0	1	3	100	0	4	1

Figure 6.11. Reader 5's evaluation of lung X-Ray images on VR and MM.

Table 6.6. Statistical comparison of Reader 5's evaluation on VR and MM.

	Reader 5-VR	**Reader 5- MM**
Number of Cases	100	100
Number of Correct	68	71
Accuracy	68%	71%
Sensitivity	68%	70%
Specificity	68%	72%
Positive Cases Missed	16	15
Negative Cases Missed	16	14
Fitted ROC Area	0.728	0.769
Empiric ROC Area	0.696	0.729
Std. Dev. (Area)	0.060	0.056

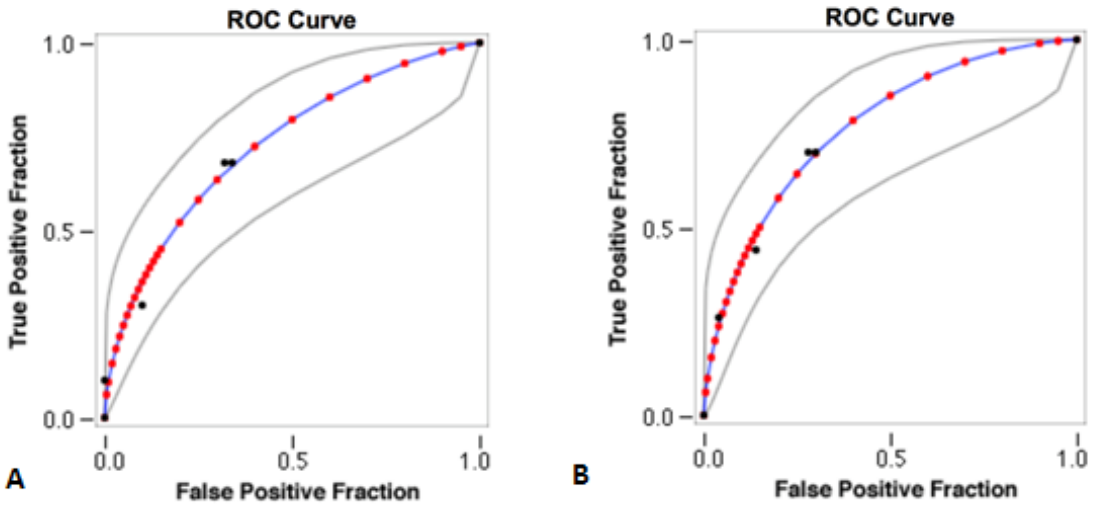


Figure 6.12. A: ROC curve of VR and B: ROC curve of MM for the Reader 5.

Table 6.7. Area under the ROC curve of the five readers for the diagnosis of X-ray lung nodules on VR and MM.

READER	**AUC for VR**	**AUC for MM**	**Difference between areas**	**P Value**
1	0.931	0.948	-0.017	0.62
2	0.865	0.915	-0.050	0.30
3	0.805	0.897	-0.092	0.09
4	0.775	0.771	0.004	0.94
5	0.728	0.769	-0.041	0.56
OVERALL	0.814	0.861	-0.047	0.06

Table 6.8. Sensitivity and specificity for the diagnosis of X-ray lung nodules using VR and MM.

Reader	Measure	VR	MM	P value
1	Sensitivity	0.76	0.82	0.58
	Specificity	0.90	0.86	0.73
2	Sensitivity	0.76	0.82	0.58
	Specificity	0.80	0.86	0.51
3	Sensitivity	0.72	0.76	0.77
	Specificity	0.88	0.78	0.23
4	Sensitivity	0.76	0.74	1.00
	Specificity	0.68	0.76	0.45
5	Sensitivity	0.68	0.70	1.00
	Specificity	0.68	0.72	0.83
OVERALL	Sensitivity	0.74	0.77	0.37
	Specificity	0.79	0.80	0.90

Table 6.9. Reviewers' (n=5) overall sensitivity for 50 nodule cases.

		Virtual Reality		P Value
		TP	FN	
MM	TP	157	35	0.37
	FN	27	31	

Table 6.10. Reviewers' (n=5) overall specificity for 50 non-nodule cases.

		Virtual Reality		P Value
		FP	TN	
MM	FP	19	32	0.90
	TN	34	165	

Table 6.11. Intra-observer agreement with VR and MM for all readers.

READER	WEIGHTED KAPPA VALUES FOR VR AND MM	STANDARD ERROR	0.95 Confidence Interval	
			Lower Limit	Upper Limit
1	0.656	0.049	0.559	0.753
2	0.782	0.048	0.687	0.877
3	0.557	0.051	0.456	0.658
4	0.383	0.051	0.281	0.485
5	0.224	0.078	0.071	0.377
OVERALL	0.542	0.026	0.491	0.594

Table 6.7 shows the Area Under Curve of all readers for the diagnosis of X-ray lung nodules on the VR and the medical monitor.

Table 6.8 shows the overall sensitivity and specificity of all readers for the diagnosis of X-ray lung nodules using VR and medical monitor.

Table 6.9 shows the overall sensitivity for 50 nodule cases.

Table 6.10 shows the overall specificity for 50 non-nodule cases.

Table 6.11 shows the intra-observer agreement with VR and medical monitor for all readers.

The evaluation time for examining 25 X-ray images on VR display is higher than medical monitor. The time difference is 2.33 minutes which is normal because of the unfamiliarity of the new technology.

For Reader 1's evaluation, AUC value for VR is less than medical monitor. Sensitivity and specificity of p value were calculated as 0.58, 0.73 respectively . p values are greater than 0.05, which indicates that two diagnostic techniques are not different for the first radiologist. Weighted kappa value was calculated 0.656 that shows there is a substantial agreement between VR and medical monitor for Reader 1.

For Reader 2, although AUC value for VR is less than medical monitor, sensitivity and specificity of p value were calculated as 0.58 and 0.51 respectively. Since the p values are greater than 0.05, this confirms that there is no difference between VR and medical monitor for Reader 2. Likewise, weighted kappa value is 0.782 that indicates the substantial agreement between two diagnostic techniques for Reader 2.

For Radiologist 3's interpretations, although AUC value for VR is less than medical monitor, sensitivity and specificity of p value were calculated as 0.77 and 0.23 respectively. Since the p values are greater than 0.05, this confirms that there is no dif-

ference between VR and medical monitor for Reader 3. Weighted kappa value is 0.557 that specifies the moderate agreement between two diagnostic techniques for Reader 3.

For Radiologist 4, AUC value for VR is greater than medical monitor, sensitivity and specificity of P value were calculated as 1.00 and 0.45 respectively. Since the P values are greater than 0.05, this confirms that there is no difference between VR and medical monitor for Reader 4. Weighted kappa value is 0.383 that gives the fair agreement between two diagnostic techniques for Reader 4. So, VR is a proper diagnostic technique for Reader 4.

For Radiologist 5's readings, AUC value for VR is slightly less than medical monitor, sensitivity and specificity of p value were calculated as 1.00 and 0.83 respectively. Weighted kappa value is 0.224 that specifies the fair agreement between two diagnostic techniques for Reader 5.

Overall, although the Area Under Curve (AUC) values of Oculus Rift CV1 are less than medical monitor, p values of McNemar test confirm that there is no difference between two techniques. Besides, weighted kappa values specify the moderate agreement between two diagnostic techniques. So, Oculus Rift CV1 is very promising while evaluating lung X-ray with nodule.

6.1.3. Experiment Discussion

In this work, we examined immersive virtual reality devices especially those targeted towards the consumer market with an emphasis on their use in tele-radiology applications where ambient light cannot be controlled and portability is required. VR is an emerging technology that is likely to bring many benefits to medicine and a powerful solution for tele-radiology. However it is important to test the potential use of such a technology in medical imaging. In particular, testing this new technology in the context of difficult reading tasks such as projection lung x-ray reading is important in order to establish new uses of VR in radiology. VR also offers advantages such as a

scalable virtual screen, decreased usage of space, low cost. We developed a VR environment proposed to perform as a virtual radiology reading room that allowed users to load and manipulate medical images. The use of a virtual environment enables users to take advantage of a wide virtual viewing area that may have been unfeasible to replicate in physical reality. In this study several analysis methods were used to assess the performance of VR compared to a standard medical monitor in a low ambient light room. These included ROC, Mc-Nemar tests and Kappa coefficients for assessing intra-observer agreement between VR and the medical monitor.

McNemar test showed that the success rate of both display systems were comparable. The areas under the ROC curves for each reader were almost similar. The comparison of the ROC results showed that the diagnostic performance of the VR was similar to medical monitor in this study.

Although portable technologies perform well for 3-D images such as CT and MRI, it would be more challenging to use them for high detail and low contrast images in daylight conditions. This result may open up new possibilities for radiological image reading particularly in tele-radiology settings. The successful test results for VR imply that this technology can be used not only for 3-D imaging tasks such as CT, MRI and PET but also for high detail and low contrast reading tasks such as nodule detection in lung x-ray images. Literature indicated that most consumer display monitors could be equivalent to medical monitors. However, the ambient light conditions may affect the performance. In the case of VR, this does not pose any difficulty. Therefore, one may conclude that VR can be an important telemedicine technology in the future, given that it can deliver high standards at variable environmental conditions and portability. In addition to the AUC and McNemar results, the Kappa coefficients indicated relatively high intra-observer agreement between VR and the medical monitor. Since Kappa values of more than 0.4 have clinical acceptance, our intra-observer agreements between the VR and LCD monitor (kappa value 0,224–0,782) can be considered acceptable. The reading time difference between VR and medical monitor was calculated to be 140 seconds. It took longer to evaluate with VR. This difference may result from the unfamiliarity with the VR technology. Some radiologists lost extra time while

calibrating, zooming, window leveling and panning the images. In future work, we would do the experiment with the remote controllers that are compatible with VR in order to decrease the assessment time of chest x-ray images. If the lesions are located in the blind spots of the lung such as apices, retrocardiac or hilar regions, they are more likely to be undetected by the readers because of the anatomic noise that is caused by the overlapping of anatomical structures of chest on each other on the radiographs. It is worth to mention that some nodules were detected by VR but not by medical monitor as shown in Figure 6.13 and 6.14. On the other hand, some nodules were detected by using medical monitor but missed by VR as shown in Figure 6.15 and 6.16. Also, Figure 6.17 shows the sample missed case with both VR and medical monitor.

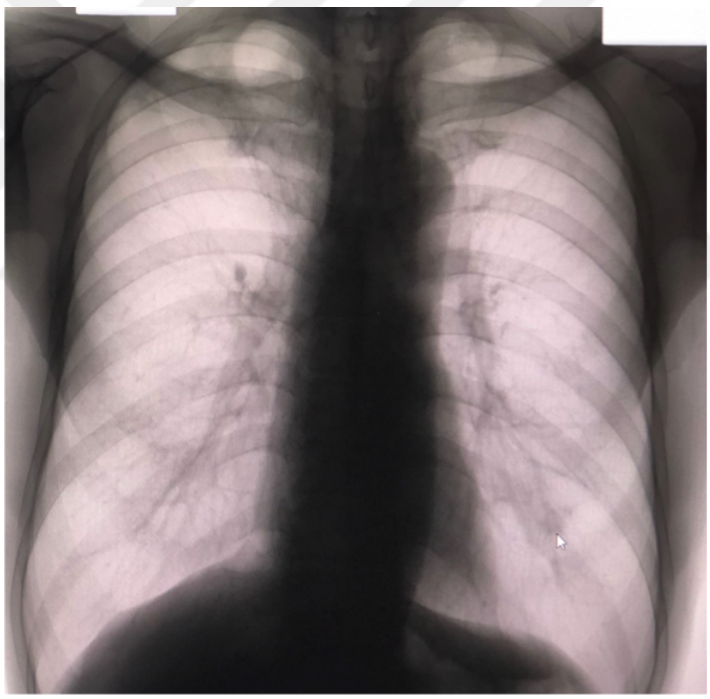


Figure 6.13. Sample missed case with MM.

The participants did not report any motion sickness after wearing VR head set. One radiologist reported discomfort after prolonged use of the Oculus which we believe this issue can be resolved in next generation VR devices. One participant praised the dark environment of the VR system noted that this specification provided focused and non-disturbed working environment for them. Some participants complained about the discomfort with the weight of the head mounted VR display and the demand of

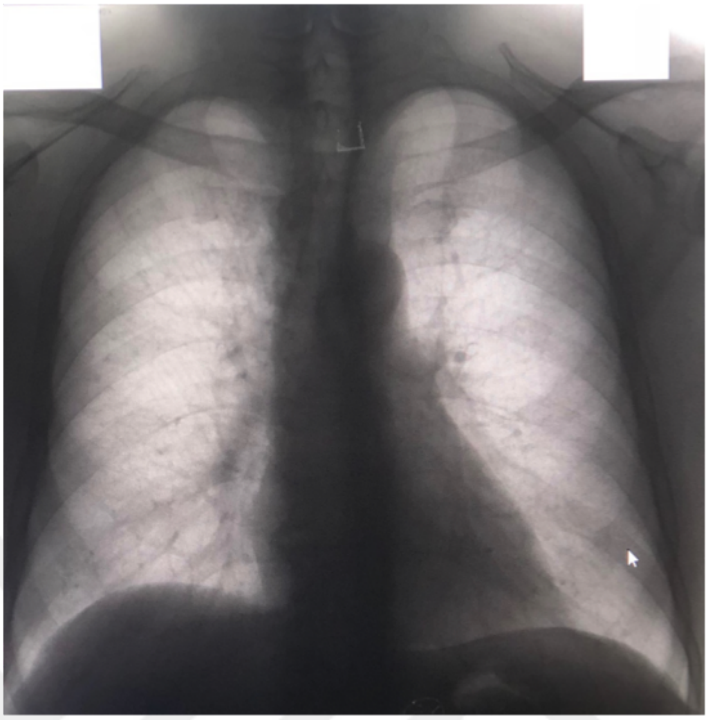


Figure 6.14. Sample missed case with MM.

adjusting the head mounted VR physically until the image was in focus. Some participants also complained about the display resolution. The results of the experiment suggest that certain radiological tasks can be achieved equally by using a VR system compared to medical monitor. Radiologists can evaluate the X-ray images even under sun light with VR which is very appealing for teleradiology.

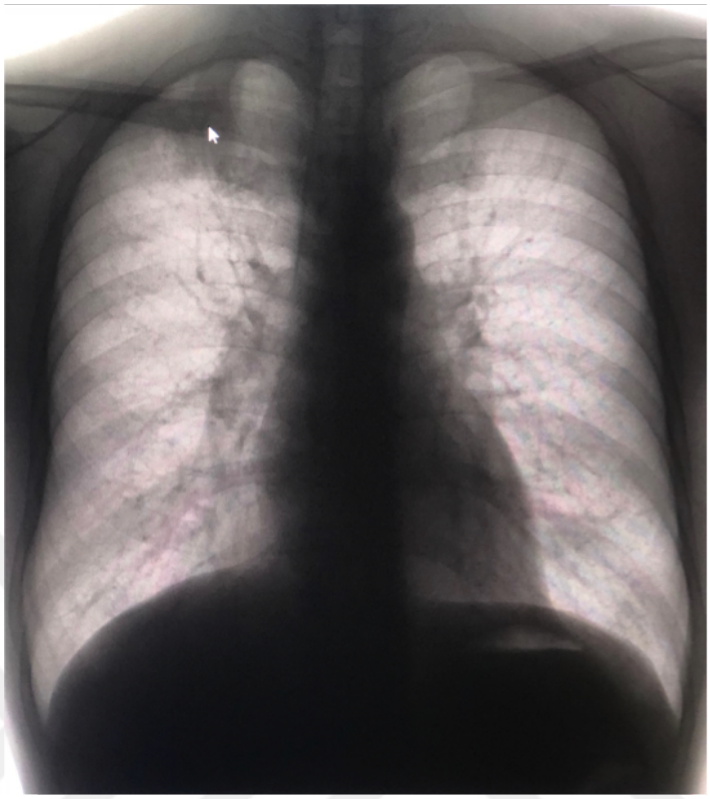


Figure 6.15. Sample missed case with VR.

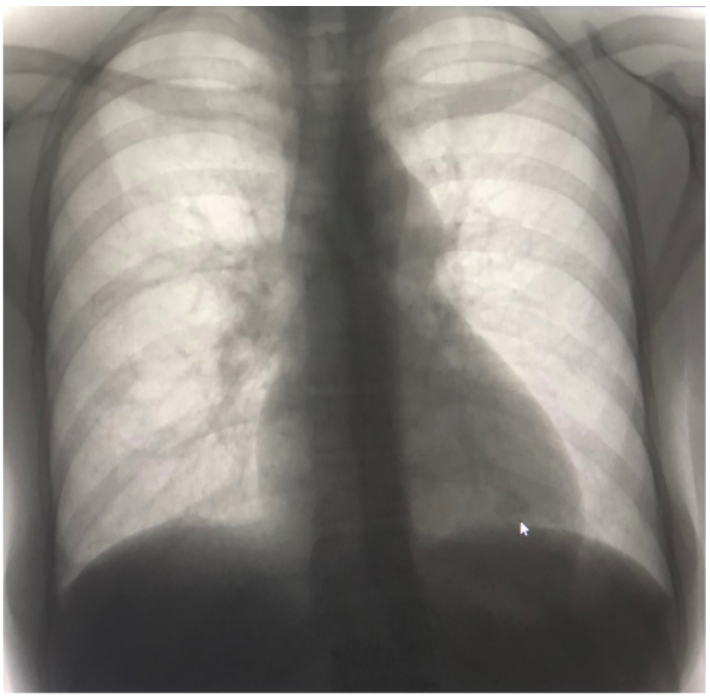


Figure 6.16. Sample missed case with VR.

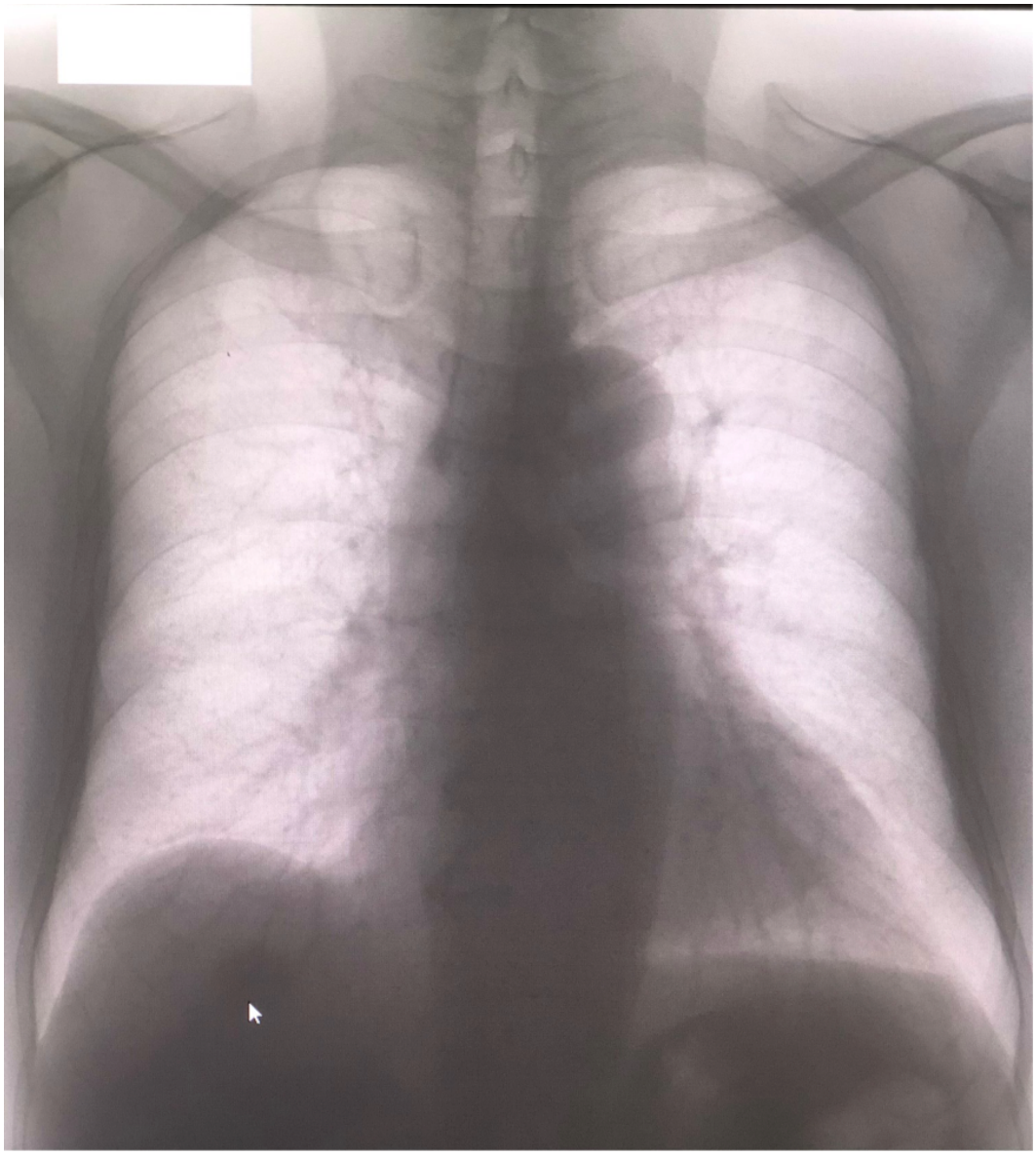


Figure 6.17. Sample missed case with both VR and MM.

7. CONCLUSION

In this study, we examined captivating virtual reality devices that primarily aims to be used in entertainment area can be a useful tool for teleradiology. Some certain advantages of Virtual Reality are offering diagnostic imaging such as scalable virtual screen, remote collaboration potential, mobility, decreased usage of space and cost in hospital settings.

We created a Virtual Reality environment that is turned into a virtual radiology reading room that allowed users (radiologists) to load and manipulate medical X-ray images outside of the hospital settings. Users had an ability to be in the environment that may have been impractical in physical reality. So, by using the VR as a diagnostic display, physicians would be free to choose the working area. Radiologists would not need to stay the dark reading room all day. Based on the feedback acquired from the experiment participants (radiologists), the Oculus Rift CV1 display has provided promising image quality to perform a differential diagnosis task with reviewed dataset.

Moreover, the ability of enrolled participants to use the virtual reality diagnostic imaging environment to accomplish radiological tasks were evaluated. The capability to detect lung nodules correctly on dataset, and the overall physical and mental requirement stated by the participants were considered.

In near future, when the next generation VR devices with better display resolution and lighter weight setup can be integrated to smart phones or tablets, the usage of VR devices in teleradiology would be widespread significantly. Even, Oculus recently announced that the Oculus Rift next version which is called Consumer version 2 with 4K resolution (6.4 times better than Oculus Rift CV1 Display) will be available in 2019.

REFERENCES

1. King, F., J. Jayender, S. K. Bhagavatula, P. B. Shyn, S. Pieper, T. Kapur, A. Lasso and G. Fichtinger, "An immersive virtual reality environment for diagnostic imaging", *Journal of Medical Robotics Research*, Vol. 1, No. 01, p. 1640003, 2016.

2. Seibert, J. A. and J. M. Boone, "X-ray imaging physics for nuclear medicine technologists. Part 2: X-ray interactions and image formation", *Journal of nuclear medicine technology*, Vol. 33, No. 1, pp. 3–18, 2005.

3. Seibert, J. A., "X-ray imaging physics for nuclear medicine technologists. Part 1: Basic principles of x-ray production", *Journal of nuclear medicine technology*, Vol. 32, No. 3, pp. 139–147, 2004.

4. *Digital Imaging and Communications in Medicine, Part 14: Grayscale Standard Display Function*, National Electrical Manufacturer's Association, 2001.

5. Milgram, P., H. Takemura, A. Utsumi and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum", *Photonics for industrial applications*, pp. 282–292, International Society for Optics and Photonics, 1995.

6. Sutherland, I. E., "A head-mounted three dimensional display", *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, part I*, pp. 757–764, ACM, 1968.

7. Pieper, S. D., M. McKenna, D. Chen and I. E. McDowall, "Computer animation for minimally invasive surgery: computer system requirements and preferred implementations", *Stereoscopic Displays and Virtual Reality Systems*, Vol. 2177, pp. 401–409, International Society for Optics and Photonics, 1994.

8. Calvano, C. J., M. E. Moran, L. D. Tackett, P. P. Reddy, K. E. Boyle and M. M. Pankratov, "New visualization techniques for in utero surgery: amnioscopy with

a three-dimensional head-mounted display and a computer-controlled endoscope", *Journal of endourology*, Vol. 12, No. 5, pp. 407–410, 1998.

9. Krapichler, C., M. Haubner, R. Engelbrecht and K.-H. Englmeier, "VR interaction techniques for medical imaging applications", *Computer methods and programs in biomedicine*, Vol. 56, No. 1, pp. 65–74, 1998.

10. Oculus, "Oculus", `http://www.oculus.com"`, accessed at May 2017.

11. Nagy, P., E. Siegel, T. Hanson, L. Kreiner, K. Johnson and B. Reiner, "PACS reading room design", *Seminars in roentgenology*, Vol. 38, pp. 244–255, WB Saunders, 2003.

12. Johnson, A. J., "From traditional reading rooms to a soft copy environment: radiologist satisfaction survey", *Journal of digital imaging*, Vol. 16, No. 3, pp. 262–269, 2003.

13. Siegel, E., B. Reiner, M. Abiri, A. Chacko, R. Morin, D. Ro, K. Spicer, N. Strickland and J. Young, "The filmless radiology reading room: a survey of established picture archiving and communication system sites", *Journal of digital imaging*, Vol. 13, No. 1, pp. 22–23, 2000.

14. WHO, "World Health Organization. Media Centre. Cancer Fact Sheet 297", `http://www.who.int/mediacentre/factsheets/fs297/en/`, accessed at July 2017.

15. Van Beek, E. J., S. Mirsadraee and J. T. Murchison, "Lung cancer screening: Computed tomography or chest radiographs?", *World journal of radiology*, Vol. 7, No. 8, p. 189, 2015.

16. NHS, "Information and Statistics. Lung cancer and mesothelioma.", `http://www.isdscotland.org/Health-Topics/Cancer/Cancer-Statistics/Lung Cancer and Mesothelioma/` , accessed at July 2017.

17. Siegel, R., D. Naishadham and A. Jemal, "Cancer statistics, 2013", *CA: a cancer journal for clinicians*, Vol. 63, No. 1, pp. 11–30, 2013.

18. Henschke, C. I., D. F. Yankelevitz and D. M. Libby, "Survival of patients with stage I lung cancer detected on CT screening", *New England Journal of Medicine*, Vol. 355, No. 17, pp. 1763–1771, 2006.

19. Prokop, M. and M. Galanski, *Spiral and multislice computed tomography of the body*, Thieme, 2003.

20. MacMahon, H., J. H. Austin, G. Gamsu, C. J. Herold, J. R. Jett, D. P. Naidich, E. F. Patz Jr and S. J. Swensen, "Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society", *Radiology*, Vol. 237, No. 2, pp. 395–400, 2005.

21. Miller, J. C., J.-A. O. Shepard, M. Lanuti, S. Aquino, J. H. Thrall and S. I. Lee, "Evaluating pulmonary nodules", *Journal of the American College of Radiology*, Vol. 4, No. 6, pp. 422–426, 2007.

22. Kundel, H. L., C. F. Nodine and D. Carmody, "Visual scanning, pattern recognition and decision-making in pulmonary nodule detection.", *Investigative radiology*, Vol. 13, No. 3, pp. 175–181, 1978.

23. Austin, J., B. Romney and L. Goldsmith, "Missed bronchogenic carcinoma: radiographic findings in 27 patients with a potentially resectable lesion evident in retrospect.", *Radiology*, Vol. 182, No. 1, pp. 115–122, 1992.

24. Turkington, P., N. Kennan and M. Greenstone, "Misinterpretation of the chest x ray as a factor in the delayed diagnosis of lung cancer", *Postgraduate medical journal*, Vol. 78, No. 917, pp. 158–160, 2002.

25. Brogdon, B., C. Kelsey and R. Moseley Jr, "Factors affecting perception of pulmonary lesions.", *Radiologic Clinics of North America*, Vol. 21, No. 4, pp. 633–654,

1983.

26. Manning, D., S. Barker-Mill, T. Donovan and T. Crawford, "Time-dependent observer errors in pulmonary nodule detection", *The British journal of radiology*, Vol. 79, No. 940, pp. 342–346, 2006.

27. Kundel, H. L. and P. S. La Follette Jr, "Visual search patterns and experience with radiological images", *Radiology*, Vol. 103, No. 3, pp. 523–528, 1972.

28. Del Ciello, A., P. Franchi, A. Contegiacomo, G. Cicchetti, L. Bonomo and A. R. Larici, "Missed lung cancer: when, where, and why?", *Diagnostic and Interventional Radiology*, Vol. 23, No. 2, p. 118, 2017.

29. Quekel, L. G., A. G. Kessels, R. Goei and J. M. van Engelshoven, "Miss rate of lung cancer on the chest radiograph in clinical practice", *CHEST Journal*, Vol. 115, No. 3, pp. 720–724, 1999.

30. Byers, T. E., J. E. Vena and T. F. Rzepka, "Predilection of Lung Cancer for the Upper Lobes: An Epidemiologic Inquiry 2", *Journal of the National Cancer Institute*, Vol. 72, No. 6, pp. 1271–1275, 1984.

31. Chotas, H. G. and C. E. Ravin, "Chest radiography: estimated lung volume and projected area obscured by the heart, mediastinum, and diaphragm.", *Radiology*, Vol. 193, No. 2, pp. 403–404, 1994.

32. Wu, M.-H., M. Gotway, T. Lee, M.-S. Chern, H.-C. Cheng, J.-C. Ko, M.-H. Sheu and C.-Y. Chang, "Features of non-small cell lung carcinomas overlooked at digital chest radiography", *Clinical radiology*, Vol. 63, No. 5, pp. 518–528, 2008.

33. Forrest, J. V. and S. S. Sagel, "The lateral radiograph for early diagnosis of lung cancer", *Radiology*, Vol. 131, No. 2, pp. 309–310, 1979.

34. Shea, F. J. and M. C. Ziskin, "Visual system transfer function and optimal viewing

distance for radiologists.", *Investigative radiology*, Vol. 7, No. 3, pp. 147–151, 1972.

35. Stitik, F. and M. Tockman, "Radiographic screening in the early detection of lung cancer.", *Radiologic Clinics of North America*, Vol. 16, No. 3, pp. 347–366, 1978.

36. Howarth, N. and D. Tack, "Missed lung lesions: side by side comparison of chest radiography with MDCT", *Diseases of the Chest and Heart 2015–2018*, pp. 80–87, Springer, 2015.

37. Rühl, R., M. M. Wozniak, M. Werk, F. Laurent, G. Mager, M. Montaudon, A. Pattermann, A. Scherrer, J.-P. Tasu, M. Pech *et al.*, "CsI-detector-based dual-exposure dual energy in chest radiography for lung nodule detection: results of an international multicenter trial", *European radiology*, Vol. 18, No. 9, p. 1831, 2008.

38. Kano, A., H. MacMahon, D. D. Hassell, M. L. Giger *et al.*, "Digital image subtraction of temporally sequential chest images for detection of interval change", *Medical Physics*, Vol. 21, No. 3, pp. 453–461, 1994.

39. Li, F., R. Engelmann, C. E. Metz, K. Doi and H. MacMahon, "Lung cancers missed on chest radiographs: results obtained with a commercial computer-aided detection program", *Radiology*, Vol. 246, No. 1, pp. 273–280, 2008.

40. Kakeda, S., J. Moriya, H. Sato, T. Aoki, H. Watanabe, H. Nakata, N. Oda, S. Katsuragawa, K. Yamamoto and K. Doi, "Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system", *American Journal of Roentgenology*, Vol. 182, No. 2, pp. 505–510, 2004.

41. White, C. S., T. Flukinger, J. Jeudy and J. J. Chen, "Use of a computer-aided detection system to detect missed lung cancer at chest radiography", *Radiology*, Vol. 252, No. 1, pp. 273–281, 2009.

42. Kumar, R. and A. Indrayan, "Receiver operating characteristic (ROC) curve for medical researchers", *Indian pediatrics*, Vol. 48, No. 4, pp. 277–287, 2011.

43. Vining, D. and G. W. Gladish, "Receiver operating characteristic curves: a basic understanding.", *Radiographics*, Vol. 12, No. 6, pp. 1147–1154, 1992.

44. Broemeling, L., "Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis. By KH Zou, A. Liu, AI Bandos, L. Ohno-Machado and HE Rockette. Boca Raton, Florida: Chapman & Hall/CRC Press. 2012. 245 pages. UK£ 57.99 (hardback). ISBN 978-1-43981-2228.", *Australian & New Zealand Journal of Statistics*, Vol. 54, No. 4, pp. 509–510, 2012.

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

46. Zabala-Travers, S., M. Choi, W.-C. Cheng and A. Badano, "Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images", *Medical physics*, Vol. 42, No. 6, pp. 2942–2954, 2015.

47. 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.

48. Obuchowski, N. A., "Receiver Operating Characteristic Curves and Their Use in Radiology 1", *Radiology*, Vol. 229, No. 1, pp. 3–8, 2003.

49. Park, S. H., J. M. Goo and C.-H. Jo, "Receiver operating characteristic (ROC) curve: practical review for radiologists", *Korean Journal of Radiology*, Vol. 5, No. 1, pp. 11–18, 2004.

50. Bamber, D., "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph", *Journal of mathematical psychology*, Vol. 12, No. 4, pp. 387–415, 1975.

51. Hanley, J. A. and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases.", *Radiology*, Vol. 148,

No. 3, pp. 839–843, 1983.

52. Sullivan, P. M., *The Statistical Evaluation of Medical Tests for Classification and Prediction*, New York: Oxford University Press, 2004.

53. Eng, J., "JHU ROC calculator", `http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi` , accessed at May 2017.

54. Addinsoft, "XLSTAT-Base McNemar Test", `http://www.xlstat.com`, accessed at May 2017.

55. Crewson, P. E., "Reader agreement studies", *American Journal of Roentgenology*, Vol. 184, No. 5, pp. 1391–1397, 2005.

56. Psoter, K. J., B. S. Roudsari, M. K. Dighe, M. L. Richardson, D. S. Katz and P. Bhargava, "Biostatistics primer for the radiologist", *American Journal of Roentgenology*, Vol. 202, No. 4, pp. W365–W375, 2014.

57. MedCalc, "MedCalc software", `https://www.medcalc.org`, accessed at May 2017.

58. McHugh, M. L., "Interrater reliability: the kappa statistic", *Biochemia medica*, Vol. 22, No. 3, pp. 276–282, 2012.

59. Dwyer, S. and B. Stewart, "Clinical uses of grayscale workstations", *Digital Imaging*, , No. 22, pp. 241–264, 1993.

60. 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.

61. Shiraishi, J., S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. Komatsu, M. Matsui, H. Fujita, Y. Kodera and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: receiver

operating characteristic analysis of radiologists' detection of pulmonary nodules", *American Journal of Roentgenology*, Vol. 174, No. 1, pp. 71–74, 2000.

62. NIH, "MIPAV software", `https://mipav.cit.nih.gov/`, accessed at May 2017.

63. Oculus, "Oculus software", `https://www.oculus.com/experiences/rift/911715622255585/`, accessed at May 2017.

64. MicroDicom, "Microdicom software", `https://www.microdicom.com/` , accessed at May 2017.