Introduction

This paper focuses on the visual evaluation of image quality in which human observers play an essential role. The two main approaches for evaluating image quality with human observers fall into two catagories: detection of pathology in which a search strategy is normally used and the assessment of anatomical structures' visibility. To evaluate observer performance in distinguishing pathology, Receiver Operating Characteristics (ROC) analysis is typically used; while visualisation of anatomical structures uses Visual Grading Analysis (VGA). The goal of image quality evaluation may be to achieve detailed information regarding image quality for the use of a new technique, identify another way to position the patient or perhaps another purpose of the study.

Accurate diagnosis for the patient is the main purpose of radiography. All optimization should be focused on producing images for more accurate diagnosis followed by better treatment of the individual patient. A radiographer is the patient attorney and is recognised by using the modalities to produce optimised medical images [1]. Optimised image quality is achieved when the clinical question can be answered whilst keeping patient radiation dose as low as reasonably possible. This can be done if the conversion of the X-rays into image information is performed as efficiently as possible [2]; the required image quality should relate to the clinical question and this is known as task based radiography [3,4].

Optimisation is a complicated process involving correct positioning of the patient, using optimal technical parameters, using optimised software parameters and having observers who are properly trained for the visual diagnostic task which is conducted in a suitable environment. Radiographic Image Quality (RIQ) is defined by spatial resolution, contrast resolution and noise [5], although with optimisation in mind these parameters should be fit for the clinical purpose of the image. To evaluate image quality in either the clinical setting or in a research project, a valid method should be used [6,7]. Direct determination of clinical performance can be difficult as this involves the overall value of the image to the patient's diagnosis in terms of diagnostic accuracy and eventually the value of diagnosis to treatment. Assessing clinical performance can be expensive and time-consuming. For research purposes, the quality of the results obtained will likely depend on the number of patients included, patient characteristics and the observers [2,8]. As an alternative to assessing clinical performance, image quality can be assessed as a surrogate for clinical performance and this can be achieved through task-oriented observer experiments. Such experiments are simpler in comparison to clinical performance studies.

Physical or anthropomorphic phantoms, cadavers or animal models can be used as well as living humans to evaluate image quality. Physical measures of technical phantoms are essential and are helpful in describing the performance of the imaging system in terms of image quality, but they do not relate to all components of the imaging chain [8-11].

ROC analysis and related methods (e.g. Free Response ROC (FROC) and JAFROC (Jacknife FROC)) are validated methods and are considered to be the gold standard for the visual assessment of image quality. This is because they provide an opportunity to assess the image in terms of its ability to demonstrate abnormalities [12, 13]. Where confidence is taken into account with JAFROC, the true positive fraction (TPF) and the false positive fraction (FPF) depend on the choice of the confidence level which results in a positive decision (threshold). A curve is created to illustrate the relationship between sensitivity and specificity for a full range of decision thresholds. The data are commonly summarised as the 'Area Under the ROC Curve' or AUC and is defined as the probability that a randomly selected abnormal case has a test result more indicative of abnormality than that of a randomly chosen normal case [14, 15]. The prerequisite of ROC analysis is that the true state of the images (normal/abnormal) must be known, referred to as ground truth. A lesion must be validated either by correct diagnosis in advance or by later follow-up diagnosis to ensure it is a real abnormality. As an alternative, artificial lesions could be digitally inserted in the image, although

the images will then have lower clinical fidelity. Still, ROC analysis is limited to the task given by the study setup and does not measure the ability of the imaging system to visualise other lesions that are important for the clinical value of the examination type or details regarding the visualised image quality are not given [2].

Alternatively, image quality can be assessed using VGA, which is a simpler and more intuitive method to measure image quality based on the visibility and reproduction of structures seen within the image [16-19]. The method is based on the assumption that "the visibility of [normal] anatomy is strongly correlated to the detectability of pathological structures" [11]. When presented with images the observers grade their personal impression of the visualisation of defined anatomical structures using an absolute or a relative rating scale [20]. The relatively easy setup of a VGA study makes it suitable for use in optimisation studies in a clinical environment.

The aim of this paper is to raise some the strengths and limitations of a VGA method for image quality evaluation and to outline the method from a radiographer's perspective.

Visual Grading Analysis

A VGA study includes a set of images that are graded by a number of observers. The gradings reflect the perceived image quality of specific anatomical structures within the image [21]. A VGA study will involve the use of image quality criteria in a scoring scale, the images to be evaluated, the observers, a display platform for visual analysis, a suitable environment in which the display system sits and finally appropriate treatment of the resultant data through statistical analysis. Bias can influence the VGA results so every step of the VGA method should be considered exhaustively.

Relative and absolute visual grading

There are two types of visual grading analysis – absolute and relative [9,11,17]. For absolute VGA, images are graded in isolation and with no reference image. The VGA scale is typically answered with scores from 1-3 or 1-5, where 1 is "not reproduced" or "the structure could not be discerned" to 5 (or 3) for "very well reproduced" or "the structure has a completely distinct shape". The advantage of this method is the statistical possibilities in using the data, although the limitation is the lack of a reference point for interpretation of image quality and this can lead to higher levels of intra- and inter-observer variability [9,17].

In relative VGA, or comparative grading, the visibility of structures in an "experimental image" are compared and graded against the same structures in a reference image. The reference image is usually the same for all experimental images. There should be clear justification for the selection of the reference image. The observers grade the visibility of each structure within the experimental image with a scale where "0" can imply visibility equal for a structure within the reference and experimental images; while negative or positive values imply inferior or superior visibility, when compared to the reference image [9,17]. The advantage of relative VGA is the intuitive rationale of the method, in, for example, an optimisation study (better or worse than the current technique), although the limitation will be the statistical possibilities of the data and the fact that all VGA scores depend on the reference image.

Image criteria

In 1996 image quality criteria for diagnostic radiographic images was developed by a group of expert radiologists and medical physicists; the outcome was published as European Guidelines for image quality [22-24]. These criteria serve to ensure the optimisation of image quality for specific examinations in adult and paediatric radiography as well as computed tomography (CT). The criteria are historic, being valid in an era of film, but they have been adopted and adapted for use in digital radiography [9,19,26]. Sund et al.

concluded that “the modified European quality criteria are useful for separating digital images with different image qualities” [11]. A significant limitation of the 1996 criteria is that they were never validated in a formal experimental setting.

VGA should use validated image criteria. VGA criteria wording should be clear and the meaning should be unambiguous to observers, see Table 1. With task-based radiography in mind the 1996 criteria have value for assessing clinical images, however the criteria have to be adjusted for individual projects before being used in VGA. Some of the 1996 criteria do not fit with the use of anthropomorphic phantoms, cadavers or animal models. For some examinations, such as cardiac CT angiography, no guidelines for image criteria are currently available. In such cases, image criteria have to be defined in close cooperation with those who are experienced in image interpretation, such as radiologists and reporting radiographers.

From a radiographic perspective we aim to discuss RIQ [5] based on VGA results. Hereby the image criteria should present the relevant parameters from RIQ as spatial resolution, contrast resolution and noise, which could be visualised in a table (see Table 2) like that illustrated by Precht et al. [27]

Term	Definition
Visualization	Characteristic features are detectable but details are not fully reproduced: features are just visible
Reproduction	Details of anatomical structures are visible but not necessarily clearly defined: detail is emerging
Visually Sharp Reproduction	Anatomical details are clearly defined: details are clear
Important Image Details	These define the minimum limiting dimensions in the image at which specific or abnormal anatomical details should be recognized

Table 1: Definition of the Degree of Visibility for Anatomical Structures in an Image [8,22-24].

No.	Image criteria	Relation to technical image quality
1	Sharp/clear demarcation of the aortic wall	Sharpness of the edge in a large structure
2	Sharpness of the coronary artery contour	Sharpness of the edge in a relatively small structure
3	Sharp/clear reproduction of the anterior mitral valve	High contrast spatial resolution
4	Homogeneity in the left/right ventricle	Noise
5	Visualization of the myocardial septum between the right and left ventricle	Low contrast resolution

Table 2: Example VGA image quality criteria connected to RIQ.

The number of image quality criteria included depends on the specific task the observer must undertake and/or the research question. If more criteria and/or images are included then more time will be needed for the observers to conduct the study, and thus the cost of the study. Additionally, increasing observer time may exclude some observers from participating because their time might be limited.

To ensure the images are acceptable for clinical use, often a final criterion is included, for example "Would this image be acceptable for diagnostic purposes?" The benefit of including this criterion is to make it possible to consider task-based radiography. Often it has been shown that an image with a relatively low VGA grading could still be approved for diagnostic use [25,27].

Images

The number of images included in a VGA study depends on the study purpose and in turn this depends upon the study design and what is being tested. First of all, the type of object to be imaged will influence the number of images. If the images have little or no variety of anatomy or body size, i.e. images of anthropomorphic phantoms, cadavers or animal models, then only the technique explored will influence the image quality and the researcher can control all parameters. If the study consists of patient images, more images are likely needed as the patient population will vary. If the difference in image quality between two imaging techniques is large, then fewer images are needed. A power calculation should be used to ensure enough images and observers are included for a meaningful outcome to be achieved [13].

Observers

Variability between and within observers included in VGA is expected, which is a major challenge. Repeated evaluations of an image by an observer using the same criteria under the same conditions is necessary to assess intra-observer variability. Inter-observer variability should also be assessed using appropriate statistical tests. The question of how many observers are needed depends on many parameters, like difference in image quality between the imaging techniques and experience of the observers. In an ideal world, 10 or more experienced observers, with experience in all investigated imaging conditions, would be good. An absolute minimum of observes will be three but at least five is recommended [1,32]. Importantly, variability within and between the observers should be measured as a quality check [3,4,7,9,18,25,27].

Prior to starting a VGA study, the observers included should be trained for the VGA task and have suitable experience to meet required standard. To minimise fatigue bias, each observer should perform the image evaluation in either the morning or in the afternoon, because observer fatigue can be a problem. Consequently, the time of day and what the observer has been doing on that day need consideration. Studies conducted in the morning / earlier in the day will generally suffer less from observer fatigue. Depending upon the number of images to be assessed and the complexity of the criteria / scale it might be necessary to split image evaluation into two or more reading sessions to minimise the chance of fatigue. Consideration should also be given to whether the observer can go back and forward between images to change their VGA scores. Also, it is important to show images in a random order and not in an ascending or descending order of image quality, in order to minimise bias [30]. Observer visual acuity should be considered too – observers should have 20:20 vision as assessed by a qualified practitioner and eyesight correction should be worn if prescribed.

The professional background and/or experience of observers can be important, especially if the results are to have clinical value. If the observers are practising radiographers or radiologists, then the results will probably have ecological validity and this likely have direct value in the clinical setting. However there is a task dependency, such that if the task does not require clinical radiological background and the task simply relates to visibility of structures using a relative grading approach then so long as the observers are adequately trained and competent to do the task then the results will likely still be valid.

Image display platforms, computer monitors and environments

When setting up an observer VGA experiment it is important to try to mimic the clinical situation as much as possible. The images should be displayed on the monitors that are used for the daily diagnostic work; if this is not possible then the monitors should meet the same specification as clinical monitors and their associated quality assurance standards [34,35]. The reason for using diagnostic quality monitors is that the monitor itself should not adversely impact the results. This is particularly important if different monitors are used in the study. If one observer uses a high-quality diagnostic monitor and another uses an ordinary desktop monitor, then the results might be different for the same images, which is not acceptable [36,37]. Having said this, some studies have showed that this might not be the case; fidelity of image detail is likely to be a key factor when matching monitor specifications to the VGA task [38,39]. Ideally, all image quality evaluations should be done on the same display monitor. If this is not possible, then all monitors should have similar characteristics and be calibrated to the same specification. Calibration should be performed according to the DICOM greyscale standard display function (GSDF) [34,40].

For accuracy and time efficiency when carrying out VGA experiments, special software platforms can be invaluable, as many clinical systems typically do not have built-in modules for observer experiments. There are several characteristics that are desirable of an image display platform. The image display platform should be able to display the images in random order for each observer, and to store the gradings from each observer separately, so that intra- and inter observer variance can be analysed. The display software should have built-in functions for panning and zooming, and for setting window level and width. Furthermore, the image display platform should be flexible so that the researcher can define the VGA criteria according to what is to be evaluated in the particular study [28-30]. Finally, the image display platform should be able to export the data that was collected during the image quality evaluation in a format suitable for further analysis, for example with Visual Grading Characteristic (VGC) Analyzer [31]. There are a few image display platform available for conducting VGA studies. One of the more popular ones is ViewDEX, developed at Gothenburg University, Sweden [29,32]. Other softwares available are Sara, developed at the University Hospital in Leuven, Belgium [33], and MedXViewer, developed at the University of Surrey, UK [30].

The final matter worth noting is the ambient light conditions in the reading room where VGA is conducted. Light levels should be controlled and suitable for diagnosis; generally speaking, lighting should be dimmed and constant and extraneous light excluded.

Statistical methods in visual grading analysis

Ratings given by observers in a visual grading study are typically given on an ordinal scale where the order of the rating steps is defined, whereas the scaling distance between the steps is not (e.g. low, medium, high). As the distribution of ordinal data is unknown, it is unwise to use statistical tools where a specific distribution is presumed (parametric). The basic non-parametric method for statistical analysis of two compared groups is the Mann-Whitney U test or the Wilcoxon W test where the given ratings for the groups are ranked on one scale and the rank order sum for each group is calculated as a measure of the difference between the groups. The Mann-Whitney-Wilcoxon test is extended to be valid for more than two compared groups in the Kruskal-Wallis test. If the samples in the compared groups are dependent (matched/paired samples) the analysis is preferably made by the Wilcoxon signed-rank test [42]. However, as these methods are sensitive to the number of ties in the rank order sum [42], the relatively few scale steps normally used in visual grading will potentially lead to a decreased accuracy in the discrimination of the compared conditions.

It can be shown that a normalized Mann-Whitney-Wilcoxon test value (to a value between 0 and 1) is equal to the AUC in a ROC analysis of the same data [43]. ROC is a more specialized method for statistical decision analysis and has been established as the dominant method for image quality evaluation in diagnostic radiology. The data analysis method used in ROC analysis was therefore an inspiration in the development of a corresponding method for analysis of visual grading data presented by Båth and Månsson in VGC [17], followed by a software for statistical analysis of VGC data, VGC Analyzer [32]. VGC Analyzer uses non-parametric resampling methods for analysis of the uncertainties in the calculated VGC-value for multiple readers and multiple cases, either paired or non-paired samples.

Non-parametric methods for statistical analysis have advantages in that the results are not affected by any assumptions of underlying data distribution. A disadvantage is, however, that they cannot easily be used to handle more complex data with multiple dependencies, as is the case in multiple regression [42]. A method for using the regression tools in standard statistical software for analysis of visual grading data has been presented by Smedby and Fredriksson in Visual Grading Regression (VGR) [21]. In VGR, the effect of adjusting multiple factors affecting the diagnostic outcome can be analyzed to achieve an indication of the optimal setting for a specific diagnostic method, with the option of individual scaling and distribution defined for each factor.

The described methods for statistical analysis have their advantages and disadvantages. The basic methods are standard statistical tools, available in statistical software and accepted by general statisticians. The methods were described in the middle of the last century and are adapted for calculation by hand or simple calculators. VGRs use tools implemented in statistical software as well, although adapted to fully use the capacity of a modern computer. For a non-statistician, to handle advanced statistical software can appear difficult, but with adequate training the operator has at their disposal powerful mathematical tools. The VGC Analyzer on the other hand is a dedicated software, specially developed for statistical analysis of visual grading data. It is easy to handle for the non-statistician and free to use for non-commercial purposes. The main difference in application between VGC Analyzer and VGR is that VGC Analyzer is dedicated to give a non-parametric statistical uncertainty description of the difference between two compared imaging conditions, whereas VGR has its special skill in the analysis of multiple factors affecting the image quality, suitable in a more complex study set up.

Limitations of VGA

Aside not designing a suitable VGA study or analysing the data fairly, a possible limitation of the VGA method was recognised by Tingberg et al. [44]. Here, the observers gave the highest VGA score to the image they liked the best. Tingberg felt this could easily represent a *beauty contest* rather than focusing on the diagnostic purpose and value of the inherent image quality. Therefore, it is important to carefully design the image criteria and to perform appropriate validation of them [44]. Another pitfall seen in using the VGA method is the statistical methods used to evaluate the VGA results. As the VGA score will always be on a non-parametric ordinal scale the statistical methods are limited and if possible, validated methods as VGC or VGR could benefit the results.

Conclusion

Suitably designed VGA studies, whether relative or absolute, have value in assessing medical images for quality. If conducted well then the outcomes of such studies can have translatable value to the clinical setting. Such studies are within the reach of radiographers in research or clinical practice areas and through

use of humans, physical phantoms, anthropomorphic phantoms or cadaver the use of VGA can be a powerful tool in optimising images and as such achieving quality that is fit for purpose and at low dose.

References

1. Metsala E, Fridell K. Insights into the methodology of radiography science. Radiography 2018; 24:105-108.
2. Verdun FR, Racine D, Ott JG, Tapiovaara MJ, Toroi P, Bochud FO, Veldkamp WJH, Schegerer A, Bouwman RW, Hernandez Giron I, Marshall NW, Edyvean S. Image quality in CT: From physical measurements to model observers. Physica Medica 2015;31:823-843.
3. Lin Y, Luo H, Dobbins JT, McAdams HP, Wand X, Samei E. An image-based technique to assess the perceptual quality of clinical chest radiographs. Med. Phys. 2012;39(11):7019-31.
4. Precht H, Gerke O, Rosendahl K, Tingberg A, Waaler D. Large dose reduction by optimization of multifrequency processing software in digital radiography at follow-up examinations of the pediatric femur. Pediatr Radiol 2014;44:239–240.
5. Bushong SC. Radiologic Science for Technologists - Physics, Biology, and Protection. 2017 (9) Canada.
6. Swensson RG. Unfied measurements of observer performance in detecting and localizing target objects on images. Med Phys 1996;23(10):1709-25.
7. Manning DJ. Evaluation of diagnostic performance in Radiography. Radiography 1998;4:49-60.
8. Seeram E, Davidson R, Bushong S, Swan H. Image Quality Assessment Tools for Radiation Dose Optimization in Digital Radiography: An Overview. Radiologic technology 2014;85(5):555-562.
9. Månsson LG: Methods for evaluation of image quality: a review. Rat. Prot. Dosimetry 2000;90(1-2):89-99.
10. Båth M. Evaluating imaging systems: practical applications. Radiat Prot Dosimetry 2010;139(1-3):26-36.
11. Sund P, Håth M, Kheddache S, Månsson LG. Comparison of visual grading analysis and determination of detective quantum efficiency for evaluating system performance in digital chest radiography. Euro Radiol. 2004;14(1):48-58.
12. Thompson JD, Manning DJ, Hogg P. Analysing data from observer studies in medical imaging research: An introductory guide to free-response techniques. Radiography 2014;20:295-299.
13. Chakraborty DP, Validation and statistical power comparison of methods for analyzing free-response observer performance studies. Acad Radiol 2008;15(12):1554-1566.
14. Hanley JA, McNeil BJ. The meaning and use of area under a receiver operating characteristic curve. Radiology 1982;143:29-36.
15. Obuchowski NA1, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Phys Med Biol. 2018;29;63(7):07TR01.
16. Lanca L, Silva A. Digital radiography detectors – a technical overview: part 2. Radiography 2009;15(2): 134-138.
17. Båth M, Månsson LG. Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. Br J Radiol. 2007;80(951):169-76.
18. Lanhede B, Båth M, Kheddache S, Sund P, Björneld L, Widell M, Almén A, Besjakov J, Mattsson S, Tingberg A, Herrmann C, Panzer W, Zankl M, Månsson LG. The influence of different technique factors on image quality of chest radiographs as evaluated by modified CEC image quality criteria. Br J Radiol. 2002;75(889):38-49.
19. Tingberg A, Sjöström D. Optimisation of image plate radiography with respect to tube voltage. Radiat Prot Dosimetry 2005;114(1-3):286-293.

20. Mraity H, England A, Hogg P. Review article: Developing and validating a psychometric scale for image quality assessment. Radiography 2014;20:306-311.
21. Örjan Smedby and Mats Fredrikson, Visual grading regression: analysing data from visual grading experiments with regression models. Br J Radiol. 2010;83(993):767-75.
22. European Commission. European guidelines on quality criteria for diagnostic radiographic images. Bruxelles, Luxemburg, 1996.
23. European Commission. European guidelines on quality criteria for diagnostic radiographic images in pediatrics. Luxemborg, 1996.
24. European Commission. European guidelines on quality criteria for computed tomography. Available at: http://www.drs.dk/guidelines/ct/quality/htmlindex.htm (accessed 18 April 2019).
25. Precht H, Waaler D, Outzen CB, Brock Thorsen JB, Steen T, Hellfritzsch MB, Aagesen H, Holst AK, Le P, Lindequist S, Rasmussen L, Tingberg A. Does software optimization influence the radiologists' perception in low dose paediatric pelvic examinations? Radiography 2019;25(2):143-147.
26. Busch HP. Image quality and dose management for digital projection radiology – final report. In: DIMOND III. European Commission. Suomenrontgenhoitajaliitto.fi/doc/diamond_III.pdf.
27. Precht H, Thygesen J, Gerke O, Egstrup K, Waaler D, Lambrechtsen J. Influence of adaptive statistical iterative reconstruction algorithm on image quality in coronary computed tomography angiography. Acta Radiologica Open 2016;5(12):1–9.
28. Landre V, Pedersen M, Waaler D. Memory effects in subjective quality assessment of X-ray images. Scandinavian Conference on Image Analysis, SCIA 2017: Image Analysis pp 314-325.
29. Svahn T, Tingberg A. Observer Experiments with Tomosynthesis. In: Reiser I, Glick S, editors. Tomosynthesis Imaging: Taylor & Francis Books, Inc.; 2014.
30. Håkansson M, Svensson S, Zachrisson S, Svalkvist A, Båth M, Månsson LG. VIEWDEX: an efficient and easy-to-use software for observer performance studies. Radiat Prot Dosimetry 2010;139(1-3):42-51.
31. Looney PT, Young KC, Halling-Brown MD. Medxviewer: Providing a Web-Enabled Workstation Environment for Collaborative and Remote Medical Imaging Viewing, Perception Studies and Reader Training. Radiat Prot Dosimetry 2016;169(1-4):32-7.
32. Båth M, Hansson J. VGC Analyzer: A Software for Statistical Analysis of Fully Crossed Multiple-Reader Multiple-Case Visual Grading Characteristics Studies. Radiat Prot Dosimetry. 2016;169(1-4):46-53.
33. Börjesson S, Håkansson M, Båth M, Kheddache S, Svensson S, Tingberg A, et al. A software tool for increased efficiency in observer performance studies in radiology. Radiat Prot Dosimetry. 2005;114(1-3):45-52.
34. Jacobs J, Zanca F, Bosmans H. A novel platform to simplify human observer performance experiments in clinical reading environments. In: Manning DJ, Abbey CK, editors. Proceedings of SPIE; Orlando, FL, USA: SPIE Press; 2011. p. 79660B-1 - 9.
35. Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, et al. Assessment of display performance for medical imaging systems. Madison, WI, USA.; 2005 April 2005. Report No.: AAPM on-line report 03.
36. Tingberg A. Suspension criteria for image monitors and viewing boxes. Radiat Prot Dosimetry. 2013;153(2):230-5.
37. Buls N, Shabana W, Verbeek P, Pevenage P, De Mey J. Influence of display quality on radiologists' performance in the detection of lung nodules on radiographs. Br J Radiol. 2007;80(957):738-43.
38. Ferranti C, Primolevo A, Cartia F, Cavatorta C, Ciniselli CM, Lualdi M, et al. How Does the Display Luminance Level Affect Detectability of Breast Microcalcifications and Spiculated Lesions in Digital Breast Tomosynthesis (DBT) Images? Acad Radiol. 2017;24(7):795-801.

39. Salazar AJ, Aguirre DA, Ocampo J, Camacho JC, Diaz XA. DICOM gray-scale standard display function: clinical diagnostic accuracy of chest radiography in medical-grade gray-scale and consumer-grade color displays. AJR Am J Roentgenol. 2014;202(6):1272-80.
40. McIlgorm DJ, McNulty JP. DICOM part 14: GSDF-calibrated medical grade monitor vs a DICOM part 14: GSDF-calibrated "commercial off-the-shelf" (COTS) monitor for viewing 8-bit dental images. Dentomaxillofac Radiol. 2015;44(3):20140148.
41. DICOM NEMA. Digital Imaging and Communications in Medicine (DICOM), Part 141999.
42. Machin D. Medical statistics a textbook for the health sciences. 4th ed. Campbell MJ, Walters SJ, editors. Chichester: Wiley; 2007.
43. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
44. Tingberg A, Båth M, Håkansson M, Medin J, Besjakov J, Sandborg M, et al. Evaluation of image quality of lumbar spine images: a comparison between FFE and VGA. Radiat Prot Dosimetry 2005;114(1-3):53-61.