

MIDA

User Testing Guide

AI-Assisted Introduction

Francisco Maria Calisto
francisco.calisto@tecnico.ulisboa.pt

04/10/2018

Prototype: [prototype-multi-modality-assistant](#) **Version:** [v1.0.1-alpha](#)
Milestone: [1.0.1-alpha](#) **Release:** [v1.0.1-alpha](#)

DICOM: [dicom-server](#)
Commit: [80191e9941c24043c7f612b2dadcd415c060bf96](#)

Deployment Environment: Localhost **Deployment Server:** Localhost
Link: [breastscreening.io/dashboard](#)

Main Server: [Localhost](#)
Private IP: [localhost](#)
Private Domain: [localhost](#)

Port: 8486
Public IP: [localhost](#)

DICOM Server: [Localhost](#)
Private IP: [localhost](#)
Private Domain: [localhost](#)

Port: 8448
Public IP: [localhost](#)
From: 8448

1 Introduction

This document aims to describe the protocol performing a set of tests in the scope of v1.0.1-alpha version from the [prototype-multi-modality-assistant](#) repository of the [MIDA](#) project using traditional devices (mouse and keyboard). The goal of the test is to understand the user, performance, efficiency and efficacy metrics in a context of an Artificial Intelligence (AI) diagnosis [7], hereby denominated as **Assistant**. The sessions will be recorded via video on the computer and using a record, heat-map and triggered event tools. It is guaranteed the confidentiality of the recordings, which will be used only for academic purposes.

Dividing the activity session into four distinct phases per each three activities representing three different scenarios (Single-Modality vs Multi-Modality vs Assistant). The first three phases took place on an [early](#) stage of the *User Tests* while we were focus to publish on the [CHI'19 Conference](#). The fourth phase will cover the hereby *User Testing Guide*. Still, we will describe as follows the overall of the four phases to give higher context.

Each scenario will have three patients. In both two scenarios, by supporting our traditional devices, the interaction is made with mouse and keyboard. The first phase, is the [demographic questionnaire](#), where we characterise the Radiologist profile. The second phase is the act of classifying those patients. Radiologists will classify each patient by using the [BIRADS](#) [1]. On the third phase, we will do several small questionnaires at the end of each scenario using [NASA-TLX](#) [18], [System Usability Scale \(SUS\)](#) [17] and measuring the [BIRADS](#) [1]. Finally, the fourth phase, and most importantly to this *User Testing Guide*, we will provide to the clinicians the **Assistant** suggestion for the [BIRADS](#) [1] results and justification of the prognostic. For the user tests we used a three distinct prototype repositories [prototype-multi-single](#), [prototype-multi-modality](#) and [prototype-multi-modality-assistant](#), the three are similar mirrors of the [prototype-breast-cancer](#) with minor changes.

2 Material

For the material and apparatus, it is essential to capture the session apprehending the user interactions. In our case, we will record this interaction by using the [QuickTime Player Version 10.4 \(928.5.1\)](#) to obtain all interactions. We will pair this video tool with a user watch tool called [Hotjar](#). This tool serves the purpose of using several logs of the interaction and gives us visualisation over it. Both instruments will help us to capture where are users interacting. By looking at the test participant's reactions, we find a lot of information regarding the prototype design.

The tools that we choose for the material and apparatus of this User Testing Guide are low-cost and easy to use. Our equipment is a cost-effective and, by using our laboratory materials, bringing it to the radiology room, we enable to capture not only what the user is doing on the screen, but on the body language supported by the interviews and observation.

The material used in the test sessions for the user interface consists of:

- MacBook Pro: it will allow the user to interact with the keyboard and a wireless mouse;
- Wireless Mouse: it will allow the user to interact with a mouse and will complement the keyboard;

2.1 Technical Details

To produce this traditional environment, and since we can simulate with a laptop, the mouse and keyboard interaction, we are using a Microsoft Mobile Mouse 4000 together with the [MacBook Pro](#) (Retina, 13-inch, Early 2015) with a standard integrated keyboard on the laptop.

2.2 Software

To track our user interactions across our system, we are using [Hotjar](#). This tool is an analytic package allowing us to follow our users remotely. It also provides two critical pieces of functionality, among others, that can aid in remote user testing. First of all, the heatmaps allow us to see where users are clicking, tapping and scrolling on our system. Second, it records a video playback of the entire user session. The tool shows evidence of being useful for our studies while we successfully used it in the past.

To record the task activities and the interview, we used [QuickTime](#) [19]. The [QuickTime](#) ([Apple Computer](#)) tool is available for [MacBook Pro](#) to movie, audio and screen recording. Despite of have an overall of features, we just used it for our user's screen recording. It provides this functionalities at minimum requirements and compatible to our apparatus.

3 Description

To verify our work, we identified measurable and explicit targets. By having several goals, including that a value percentage of the users should be able to operate the tasks without the need of help. On the same rate value, the user should be able to start and complete the medical diagnosis tasks over the system with little errors or mitigating those errors. Measuring the expected number of errors with a relation between our laboratory pilot tests. On the laboratory pilot tests we aim to test our prototypes with researchers. The Researchers are in the context of the system and know well the functionalities so that we need to expect a percentage value over their results compared to clinicians and not the same benefits. Last but not least, both users (researchers and clinicians) should be able to understand in a similar time amount the meaning of all visible controls. By the similar amount of time, it is expected to have a variance of the percentage value between researchers and clinicians of the same value percentage of the early goals described in this paragraph.

We tested each objective in early laboratory and field tests so that we could take the appropriate corrective actions. Also, we expect to run early field tests with researchers and clinicians to highlight issues that we overlooked and ignored during the prototyping phase. To support interaction use by the clinicians, we will try to emphasise several key factors on our user tests. The tasks must be simple, low intrusive, support for natural interaction and the system must always give visibility and the task current-state.

3.1 Devices

Traditional interaction remains the most common way to interact with user interfaces in a clinical environment. Unfortunately, most of this interaction is made by low profile equipment that makes users produce more errors and take more time interacting with those user interfaces.

On Figure 1 the user can select the list of patients. The list has a table with several patient information. The first column is the *Patient ID*; we used it as an identifier of the patient. That way we can have anonymised information with no reference to the patient name. The second column is the *Study Date*, the third column is the *Modality* of the used **DICOM** image, the fourth column is the *Study Description* of the used study and the last column is the number of *Images*.



Study List					Help	About
Patient ID	Study Date	Modality	Study Description	# Images		
202732	20180309	MAMA^ROTINA	MAMA^ROTINA	2877		
440624	20180314	01	01	8		
737037	20180308	Breast	Breast	1		
22586	20180314	01	01	8		
866141	20180314	Breast	Breast	1		
586890	20161012	Breast	Breast	1		
590463	20180315	Breast	Breast	1		
570100	20180314	Breast	Breast	1		

Figure 1: List of Patients.

As we can see in Figure 2, it shows the first task in our User Interface (UI), where the patient's breasts are on a small left column. The options are in a short row near of the viewport and described below. We also have the tabs where the user can change the patient. The centre viewport shows the **DICOM** image, and it can be configured to display a number up to four **DICOM** images at the same time. The viewport has some text information on it (yellow) with the details of the metadata. Nevertheless, the **Assistant** suggestions are shown on the top-right corner of the system.

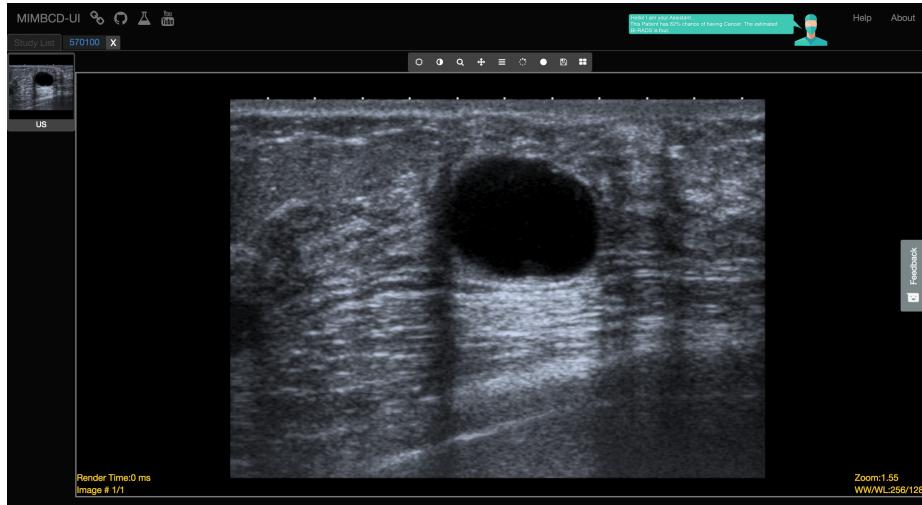


Figure 2: Viewer of the **DICOM** images.

Manual annotation is adopted by us thanks to Freehand ROI and Probe annotation features, both from [CornerstoneJS](#). According to the [CornerstoneJS](#) Library, the user can create an annotation by setting up consecutive landmarks around a Region of Interest (ROI). The markers finish a lesion annotation when it interconnects the historical. Additional features available in our User Interface (UI) includes on-demand increment of the number of landmarks, and throw transformations of the shape of an annotation.

3.2 User Interactions

The systems have several buttons (Figure 3) that allows the user to interact or access to a set of user interface features. Each item of the following list represents each metaphoric icon of Figure 3.



Figure 3: Toolbar of the System available features.

The buttons are (from left to right of Figure 3) as follows:

- WW/WC
- Invert
- Zoom
- Pan
- Stack Scroll
- Freehand
- Probe
- Save
- Window Controller

3.3 Evaluation

Introduction of AI **Assistant** agents are significant factors which can naturally affect the performance of a medical workflow. While some prior studies [2] have investigated the functionality of healthcare systems, the *AI-Assisted* acceptability has mostly been overlooked in the existing Health Informatics (HI) literature regarding a Human-Computer Interaction (HCI).

The following Table 1 is presenting seven evaluation questions to have in mind during evaluation. The purpose of this questions is to facilitate systematic user studies regarding our novel **Assistant** in a clinical environment and support user stimulation for the introduction of *AI-Assisted* methods. The proposed issues involve various aspects of workflow combined with either need for satisfaction or division of attention.

Number	Issues of Content & Key Questions
1	How would the user describe the potential adoption of <i>AI-Assisted</i> methods on the Health Institution?
2	What are the user oppositions for <i>AI-Assisted</i> methods?
3	What examples of <i>AI-Assisted</i> methods does the user know regarding the Health Institution?
4	What are the obstacles of the user's Health Institution?
5	What is more important for the <i>AI-Assisted</i> information, the BIRADS or Pathology?
6	Is it important for the user to have the feature of Approve, Reject and Justify options?
7	For the user's opinion, what are the aspects that influence the decision?

Table 1: Usability Evaluation Questions

The influence of *AI-Assisted* [10] is an important variable for our empirical analysis. In fact, the trust of the user increases when the user perceived that the **Assistant** is giving the right inputs and that there will be a consequent increase of the clinician trust in our system.

The first question, the *How would the user describe the potential adoption of AI-Assisted methods on the Health Institution?* question. For the second question, the *What are the user oppositions for AI-Assisted methods?* question, we aim to understand what are the user constraints regarding an AI adoption on the user's current workflow. Third, we intend to filter possible examples of the clinical applications of AI on the Health Institutions by asking *What examples of AI-Assisted methods does the user know regarding the Health Institution?* directly to the clinician. The fourth question, underlines the reasons why several obstacles are present on the Health Institution, with the question *What are the obstacles of the user's Health Institution?* we can understand the challenges of achieving those issues and what are the solutions for surpass it. On the fifth question, where we ask *What is more important for the AI-Assisted information, the BIRADS or Pathology?*, we aim to understand what is more important for the user, the BIRADS or the Pathology of the patient [6]. Almost last, the six question, where we ask for *Is it important for the user to have the feature of Approve, Reject and Justify options?* is an important question to understand the feature needs and options. Last but not least, the seven question, *For the user's opinion, what are the aspects that influence the decision?*, is where we will understand what is the most important information to show to the clinicians, therefore, we can more effectively and efficiently give more accurate information to the users.

To conclude this section, by doing this questions, we aim to support our user studies by giving our users, the clinicians, the opportunity of improving our empirical analysis regarding user's *open answers*. However, the results should be treated with caution. Several bias exists since we are doing here an ambiguous approach.

4 Methodology

The hereby prototype used is the [v1.0.1-alpha](#) version of our [prototype-multi-modality-assistant](#). The purpose of this prototype is to involve an *AI-Assisted* tool (**Assistant**) for medical imaging at a breast screening diagnosis level. This **Assistant** was created with a front-end and bak-end architecture utilising common programming languages, libraries, frameworks and tools including [JavaScript \(JS\)](#) [9], [NodeJS](#) [23], [HammerJS](#), [CornerstoneJS](#) [11] and [Orthonanc](#) [12]. For the Machine Learning (ML) and Deep Learning (DL) component we will use several [MATLAB](#) technologies [21], promoting and feeding our Convolutional Neural Networks (CNN) and Deep Reinforcement Learning (DRL) techniques. Other central component of this prototype is a web-based [PACS](#) [4] pairwise with ubicous web technologies and based on the [Open Source \(OS\)](#) [CornerstoneJS](#) library [8, 11].

4.1 Environments

This section describes the user environment over interaction, the so called **Radiology Room (RR)** (Figure 4). This guide is based on soft-copy diagnosis using computer workstations in their current reading room environment. It will be here where we take impressions regarding the efficacy of radiologists, and their recommendations based on their experience for improvements on the soft-copy reading environment. Several studies demonstrated [22] that radiologist fatigue levels and performance are related to environmental factors such as number of false-negative and false-positives in addition to workstation enhancements. Supported by this guide, our research aims to conduct an investigation for the several environmental variables and improvements regarding the potentially enhancement that an *AI-Assisted* diagnosis could take in the **RR**. We expect to analyse the needed information and best solution to improve the workstation results.

4.2 Case Studies

The functionality of the prototype will be best demonstrated by a series of case studies. By describing the expected workflow and capabilities of the research study at the **RR** specific environment and changes of the workflow by using our system prototype. The study implies the evaluation of medical imaging *AI-Assisted* features on several breast lesions. The primary goal of this case studies analysis is to generate a receiver operating characteristic to evaluate the performance and validation of our **Assistant**. Let us consider a list of hypothetical use cases for the research investigation that evaluates the interaction and usability performance of the **Assistant**. Therefore, the following list will show the preliminary case studies.



Figure 4: Radiology Room

List of case studies to analyse our solution prototype:

- Multi-Modality *AI-Assisted* of a Breast Cancer Diagnosis;
- Priority & Minimal Information Visualisation;
- Performance & Response Measurement Values Acquisition;
- Radiologist Validation;

We expect to demonstrate several uses through a series of case studies, including implementation of our research prototype using an *AI-Assisted* technique and features for several view studies and other imaging research, as well as creation of a novel **Assistant** for the purpose. By creating a set of questions, we will try to achieve and feed this case studies. It is automatically associated with all cases. The radiologist may interact with our **Assistant** while manipulating the medical imaging and report to us difficulties and improvements. The number of questions is not restricted to the present document, since the interview will be open and suggestive [13]. There is no limit to the number of questions that can be asked per case but it should fit the amount of expected time per each.

The data will be collect from the study video and observations into a [spreadsheet](#) for further analysis. We will [report](#) the results of this tests and conclusions. This guide and respective use cases will be iteratively improved.

5 Procedures

Participants will take part in the tests at our formed institution protocols (e.g. [Hospital Fernando Fonseca \(HFF\)](#)) with the [v1.0.1-alpha](#) version of our [prototype-multi-modality-assistant](#) repository. The interaction with the system will be used in a typical **RR** environment. Note takers and data logger(s) will monitor the sessions for observation in the **RR**, connected by screen recording feed. The test sessions will be recorded and further analysed.

5.1 Briefing

A presentation of the **Assistant** and it's use and capabilities will be made. Participants will be presented to the available interactions and will be explained how to interact with the prototype, underlining the limitations. The facilitator will brief the participants on the **Assistant** application and instruct the participant that they are evaluating the application, rather than the facilitator evaluating the participant. Participants will sign an informed consent that acknowledges: the participation is voluntary, that participation can cease at any time, and that the session will be videotaped but their privacy of identification will be granted. The facilitator will ask the participant if they have any question.

5.2 Post-Task Questionnaire

Our metrics will refer the *AI-Assisted* involvement against specific goals necessary to satisfy several requirements of our **Assistant**. For our **Post-Task Questionnaire** we will use *Quantitative Analysis* (QtA) and *Qualitative Analysis* (QIA) in response to our questions (Section 3) to measure the acceptability of our **Assistant** each time a *scenario* is completed (Section 3). From a set of tasks (Section 6) we aim to cover our main scenario, an *AI-Assisted* that we call **Assistant**. Therefore, both QtA and QIA will allow the facilitator to quickly and easily assess the requirements of the given scenario. Our QtA and QIA requirements will have several attributes [13] that make it a good choice for our clinical participants. Those attributes are as follows.

List of the scale attributes:

- The requirements are technology agnostic, making it flexible enough;
- The requirements are relatively quick and easy to answer;
- The requirements provide a single score on a scale that is easily understood;
- The requirements are nonproprietary, making it a cost effective tool;

The facilitator will explain that the amount of time taken to complete the *tasks* will be measured and that exploratory behaviour outside the *task* flow should not occur until after task completion. At the beginning of each task, the participant will listen the *task* description from the facilitator and begin the task. *Time-on-Task* (ToT) measurements begins when the participant starts the *task*, measured until the end of each *task*.

5.3 Training Session

The participant will receive and overview the test procedure. However, the user will not receive information how to annotate and interact in all degrees of freedom. With the aim of disabling users to get their work done before the test tasks. It will take advantage of a "surprise" acknowledgement.

5.4 Execution of Tasks

The *tasks* were derived from test scenarios developed from **Case Studies**. Due to the range and extent of functionality provided by our **Assistant**, and the short time from which each participant will be available, the *tasks* are the most common and relatively complex of available functions. The *tasks* are the identical for all participants of a given user role in the study.

The *tasks* will be performed by several classes of radiology experience. Professionals from Radiology Seniors, Middles, Juniors and Interns will be performing these *tasks*. On the **RR** the Radiologist is characterised [5, 15] as a physician who examines and interpret Medical Imaging (MI) [14], such as X-Rays, CT Scans or MRIs.

5.5 Post-Activity Questionnaire

After completing all *tasks* and scenarios, participants will be asked to complete a questionnaire to classify the **Assistant** according to various parameters regarding the several features. To measure this, we will use an open session *observation* and *interview* [3]. We will use this techniques to identify participants' requirements during the various stages of the workflow.

6 Tasks

During our user tests, we need to ask participants to provide a subjective assessment of their experience using our **Assistant**. There are several widely used questionnaires giving us different pros-and-cons. However, in most cases, a [single question instrument](#) [20] is the right method for a quantitative usability testing. By taking less time and effort to answer, participants are pursuing to this phase after task while it is minimally disruptive.

We will try to understand if, with the *AI-Assisted* techniques, the clinicians will encounter the most accurate severity (**BIRADS**) of the breast lesions [16] and patient's prognostic. For this purpose, we have three patients (Figure 5); each patient has three images in the respective modalities: (i) MG; (ii) US; and (iii) MRI. The clinicians will proceed to the activity of diagnosing the three patients within the support of our **Assistant** by the observation of ALL images.

In our **User Testing Guide** a set of tasks is necessary and carefully crafted. Our test studies involve asking participants to perform a set of tasks. By looking at what our user need to do with our system, our tasks are realistic as possible. We are not describing the exact steps participants need to take. We achieve that by avoiding the precise language used as labels in our system. The tasks are emotionally neutrals. And we did several **pilot tests** to prevent misleading situations saving us from wasting resources by accidentally use a lousy task or from getting bad data. The tasks are as follows.

List of stand alone tasks:

Task 1.1: Classify *Patient 1* on the **Assistant**;

Task 1.2: Classify *Patient 2* on the **Assistant**;

Task 1.3: Classify *Patient 3* on the **Assistant**;

Task 2.1: Freely explore *Patient 1* on the **Assistant**;

Task 2.2: Freely explore *Patient 2* on the **Assistant**;

Task 2.3: Freely explore *Patient 3* on the **Assistant**;

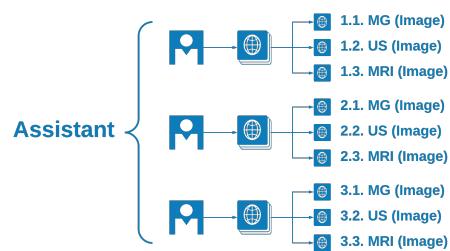


Figure 5: Diagram representing the use of the **Assistant** by clinicians.

7 Measurements

Our measurements refers to user performance measured against specific performance goals necessary to satisfy requirements. *Task* completion success rates, adherence to dialog scripts, error rates and subjective evaluations will be used. *Time-to-Completion* (TtC) of *tasks* will also be collected. The measures are as follows.

The tests are intended to achieve the following measures:

- BIRADS Classification;
- Pathology Classification;
- Time measurement;
- Number of clicks;
- Number of errors;
- Efficiency;
- Difficulty;
- Experience;

To prioritise recommendations, a method for problem difficulty and degree severity classification, as well as, pathology importance, will be used in the analysis of the collected data during evaluation process. The approach treats problem severity has a combination of several factors. Those factors are measuring the impact of the problem and the frequency of users experiencing issues during the evaluation. Nevertheless, the opinion will also be of chief importance and we will also register the received ones.

Through the questionnaire after the test session, we intend to obtain the answers to the following questions for each *task*:

- Acceptability of the **Assistant** lesion classification;
- Acceptability of the **Assistant** interaction;
- Acceptability of the **Assistant** translation;
- Acceptability of the **Assistant** available features;
- **Assistant** degrees of classification;
- **Assistant** degrees of interaction;
- **Assistant** degrees of information visualisation;

8 Goals

The next sections will describe the goals for [prototype-multi-modality-assistant](#) expectations. We will try to assess performance-related metrics such as time and correctness of participants completing *tasks* for our expectations. Our expectations are based of the [results](#) obtained at the lab as [pilot tests](#).

8.1 Completion Rate

Completion Rate is the percentage of test participants who successfully complete the task without critical errors. A critical error is defined as an error that results in an incorrect or incomplete outcome. In other words, the completion rate represents the percentage of participants who, when they are finished with the specified task, have an "output" that is correct.

A Completion Rate of 90% is the goal for each task in this usability test.

Note: If a participant requires assistance in order to achieve a correct output then the task will be scored as a critical error and the overall completion rate for the task will be affected.

8.2 Error-Free Rate

Error-Free Rate is the percentage of test participants who complete the task without any errors (critical or non-critical errors). A non-critical error is an error that would not have an impact on the final output of the task but would result in the task being completed less efficiently.

An Error-Free Rate of 80% is the goal for each task in this tests.

8.3 Time on Task (ToT)

The time to complete a scenario is referred to as "Time on Task" (ToT). It is measured from the time the participant begins the scenario to the time which the participant signals completion.

8.4 Subjective Measures

Subjective opinions about specific tasks, time to perform each task, features, and functionality will be surveyed. At the end of the test, participants will rate their satisfaction with the overall system. Combined with the interview/debriefing session, these data are used to assess attitudes of the participants.

9 Reporting Results

A [Test Report](#) will be provided at the end of this tests. It will consist of a report and/or a presentation of the results; evaluation of the metrics against the pre-approved goals, subjective evaluations, and specific issues of the system, as well as, recommendations for resolution. The recommendations will be categorically sized by development to aid in implementation strategy. The results will be translated to a [spreadsheet](#) (view only). Also, more related information can be found at [Test 4: Assistant](#).

References

- [1] Corinne Balleyguier, Salma Ayadi, Kim Van Nguyen, Daniel Vanel, Clarisse Dromain, and Robert Sigal. Birads classification in mammography. *European journal of radiology*, 61(2):192–194, 2007.
- [2] Francisco M. Calisto, Alfredo Ferreira, Jacinto C. Nascimento, and Daniel Gonçalves. Towards touch-based medical image diagnosis annotation. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, ISS ’17, pages 390–395, New York, NY, USA, 2017. ACM.
- [3] Pascale Carayon, Sarah Kianfar, Yaqiong Li, Anping Xie, Bashar Alyousef, and Abigail Wooldridge. A systematic review of mixed methods research on human factors and ergonomics in health care. *Applied ergonomics*, 51:291–321, 2015.
- [4] Robert E Cooke Jr, Michael G Gaeta, Dean M Kaufman, and John G Henrici. Picture archiving and communication system, June 3 2003. US Patent 6,574,629.
- [5] Ruth Ann Ehrlich and Dawn M Coakes. *Patient Care in Radiography-E-Book: With an Introduction to Medical Imaging*. Elsevier Health Sciences, 2016.
- [6] Eda Elverici, Ayşe Nurdan Barça, Hafize Aktaş, Arzu Özsoy, Betül Zengin, Mehtap Çavuşoğlu, and Levent Araz. Nonpalpable bi-rads 4 breast lesions: sonographic findings and pathology correlation. *Diagnostic and Interventional Radiology*, 21(3):189, 2015.
- [7] Wenjuan Fan, Jingnan Liu, Shuwan Zhu, and Panos M Pardalos. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (aimdss). *Annals of Operations Research*, pages 1–26, 2018.
- [8] Joseph Feller, Brian Fitzgerald, et al. *Understanding open source software development*. Addison-Wesley London, 2002.
- [9] David Flanagan. *JavaScript: the definitive guide. ”* O'Reilly Media, Inc.”, 2006.
- [10] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [11] Jason Hostetter, Nishanth Khanna, and Jacob C Mandell. Integration of a zero-footprint cloud-based picture archiving and communication system with customizable forms for radiology research and education. *Academic radiology*, 2018.

- [12] S. Jodogne, C. Bernard, M. Devillers, E. Lenaerts, and P. Coucke. Orthanc – A lightweight, RESTful DICOM server for healthcare and medical research. In *Biomedical Imaging (ISBI), IEEE 10th International Symposium on*, pages 190–193, San Francisco, CA, USA, April 2013.
- [13] Ger Joyce, Mariana Lilley, Trevor Barker, and Amanda Jefferies. From healthcare to human-computer interaction: Using framework analysis within qualitative inquiry. In *International Conference on Applied Human Factors and Ergonomics*, pages 93–100. Springer, 2017.
- [14] Y Kobashi, Y Munetomo, A Baba, S Yamazoe, and T Mogami. Evaluation of the ossification of the cervical poste-rior longitudinal ligament utilizing x-ray, ct and mr imaging. *Orthop Res Traumatol Open J*, 2(1):35–39, 2017.
- [15] Diana L Miglioretti, Rebecca Smith-Bindman, Linn Abraham, R James Brenner, Patricia A Carney, Erin J Aiello Bowles, Diana SM Buist, and Joann G Elmore. Radiologist characteristics associated with interpretive performance of diagnostic mammography. *Journal of the National Cancer Institute*, 99(24):1854–1863, 2007.
- [16] American College of Radiology. BI-RADS Committee. *Breast imaging reporting and data system*. American College of Radiology, 1998.
- [17] Konstantina Orfanou, Nikolaos Tseliros, and Christos Katsanos. Perceived usability evaluation of learning management systems: Empirical evaluation of the system usability scale. *The International Review of Research in Open and Distributed Learning*, 16(2), 2015.
- [18] Anjana Ramkumar, Pieter Jan Stappers, Wiro J Niessen, Sonja Adebarh, Tanja Schimek-Jasch, Ursula Nestle, and Yu Song. Using goms and nasa-tlx to evaluate human–computer interaction process in interactive segmentation. *International Journal of Human–Computer Interaction*, 33(2):123–134, 2017.
- [19] Melissa R Rowell, Frank M Corl, Pamela T Johnson, and Elliot K Fishman. Internet-based dissemination of educational audiocasts: a primer in podcasting-how to do it. *American Journal of Roentgenology*, 186(6):1792–1796, 2006.
- [20] J Sauro. 10 things to know about the single ease question (seq). *Measuring U*, 2012, 2012.
- [21] Andrea Vedaldi and Karel Lenc. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 689–692. ACM, 2015.
- [22] Stephen Waite, Srinivas Kolla, Jean Jeudy, Alan Legasto, Stephen L Macknik, Susana Martinez-Conde, Elizabeth A Krupinski, and Deborah L Reede. Tired in the reading room: the influence of fatigue in radiology. *Journal of the American College of Radiology*, 14(2):191–197, 2017.

- [23] Jim Wilson. *Node.js 8 the Right Way: Practical, Server-side Javascript that Scales*. Pragmatic Bookshelf, 2018.