

# Carnegie Mellon University in Qatar

AI for Medicine

15-182 - Spring 2023

## Assignment 1

Name: \_\_\_\_\_

Andrew ID: \_\_\_\_\_

**Due on:** January 30, 2023 by midnight

### Instructions:

- This assignment consists of three problems. Answer them all.
- Submit your answers through Gradescope.

Question	Points	Score
How to Read Articles that Use Machine Learning	30	
Self-Diagnosis Using AI	50	
Comparing Two AI Medical Diagnostic Algorithms	20	
Total:	100	

## Problem 1: How to Read Articles that Use Machine Learning (30 Points)

Here is an introductory paper that may help you start reading articles that relate to AI in general and Machine Learning (ML) in particular:

*Liu, Yun, et al. "How to read articles that use machine learning: users' guides to the medical literature." Jama 322.18 (2019): 1806-1816.*

The paper provides a high-level overview of ML and illustrates how to assess its applications in medicine. More precisely, it: (1) emphasizes the importance of proper validations of ML models in medical contexts, (2) reviews the basics of ML, and (3) suggests some considerations for clinical implementations of ML.

After reading this paper, you should be able to understand the general terminology associated with ML and identify the most crucial elements related to its validation. As we progress through the course, we will delve much deeper into many of the presented concepts, hence, do not worry much if you realize that some of these concepts are not very clear or hard to understand.

Read the paper carefully and answer the following questions:

10pts

(a) **Short Summary:**

Write a short *objective* summary of the paper in your own words (do not copy and paste). Make sure that the following questions are addressed in your summary:

- What were the paper's objectives? (1-2 sentences)
- How did the paper fulfil those objectives? (1 paragraph)
- What were the main ideas or concepts that were discussed? (2-3 paragraphs)
- What was the authors' main conclusions? (1 paragraph)

20pts

(b) **Critical Appraisal:**

Critical appraisal is a systematic evaluation of scientific research papers. In this question you will write a 2-page critical appraisal of the above Liu, Yun *et al.* paper. In your appraisal, you should identify the strengths and weaknesses of the paper, while specifically addressing the below listed questions.

We advise you to structure your appraisal around the usual chronology of the paper. The below questions are only meant to guide you while writing your appraisal; do not write your appraisal in a question/answer or a bullet-point style. Also, you do not have to limit yourself to only the below questions.

- Does the study address a clearly focused research question?
- How does the authors' specific research question fit into what is already known about the subject?
- Do the authors build a logical case and context for their hypothesis?
- Was the selected clinical scenario defined clearly?
- Was the provided guide in the paper clear and adequate to soundly resolve the selected scenario?

- Is the provided guide in the paper generalizable (i.e., applicable to any other scenario)?
- Are there any results or analyses missing that you would have wanted to see?
- What are the major weaknesses of the paper?
- What are the major strengths of the paper?
- Are the results of this paper important? What implications can they have?

## Problem 2: Self-Diagnosis Using AI (50 Points)

Read the paper “Evaluating the Accuracy of a New Artificial Intelligence Based Symptom Checker: A Clinical Vignette Study” and answer the following questions:

18pts

(a) Write a 2-page summary of the paper.

9pts

(b) Discuss at least 3 strengths of the paper.

9pts

(c) Discuss at least 3 weaknesses of the paper.

14pts

(d) How do you foresee the future of medical diagnosis as AI evolves further? Explain and justify your opinion (e.g., you can read and cite papers that discuss this topic).

## Problem 3: Comparing Two AI Medical Diagnostic Algorithms (20 Points)

In this problem you will compare the performance of two AI medical diagnostic algorithms using standard accuracy metrics.

Gold-standard					
Vignette 1		Vignette 2		Vignette 3	
Disease	Relevance	Disease	Relevance	Disease	Relevance
A	5	F	3	A	1
B	4	E	2		
C	3	C	1		
D	2				
E	1				

Table 1

<i>SuperMD</i> Results		
Vignette 1	Vignette 2	Vignette 3
B	F	A
E		

Table 2

Consider 3 gold-standard vignettes, each with a differential diagnosis as shown in Table 1. The letters, A, B, C, D, and E refer to diseases and the disease at the top of each differential list is considered as the main diagnosis in the list. Each disease in a differential list is

assigned a relevance score as illustrated in the table. The higher the relevance of a disease, the more probable it is to be the right diagnosis.

Assume there are two AI medical diagnostic algorithms named *SuperMD* and *DigitalMD*. The 3 vignettes given in Table 1 were tested on *SuperMD* and *DigitalMD*, after which results were collected and demonstrated in Table 2 and Table 3, respectively.

<i>DigitalMD</i> Results		
Vignette 1	Vignette 2	Vignette 3
E	C	A
B		

Table 3

15pts

(a) Calculate and plot in a bar chart the average M1, average M2, average M3, average recall, average precision, average F1-measure, and average NDCG of *SuperMD* and *DigitalMD*.

5pts

(b) Which algorithm, *SuperMD* or *DigitalMD*, is better in your opinion. Explain.