

Introduction to Clinical Statistics in Medical AI

Medical AI 분야 임상통계 업무 소개

서울 R Meetup 2023.06.01

곽수영

Table of Contents

- **Background**
 - Personal
 - Biostatistics
 - Medical AI
- **Main contents**
 - Biostatistical Application in Clinical Study
 - Statistical Analysis for Regulatory Approval
 - Use of AI in Work
- **Summary**
- **Q&A**

Education



Biology:
BS(학사)



Biostatistics:
MS(석사)



Background - Personal

Work



Background - Biostatistics

응용통계

- 임상통계 / 의학통계라고도 부름
- 임상 / 의학 / 생물(특히 유전) 쪽 데이터

전공과목

- Linear Regression / ANOVA / Linear Mixed Model / Survival Analysis
- Categorical Data Analysis / Nonparametric 등

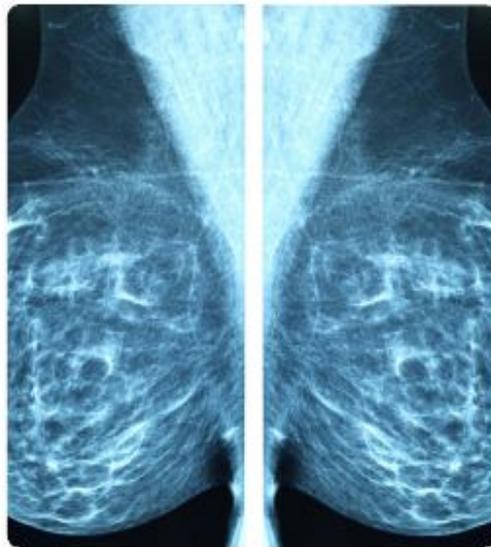
전공 후 진로?

- 병원 연구소 / 제약회사 / CRO 취직
- 국가기관(질병관리본부, 식약처 등) 취직

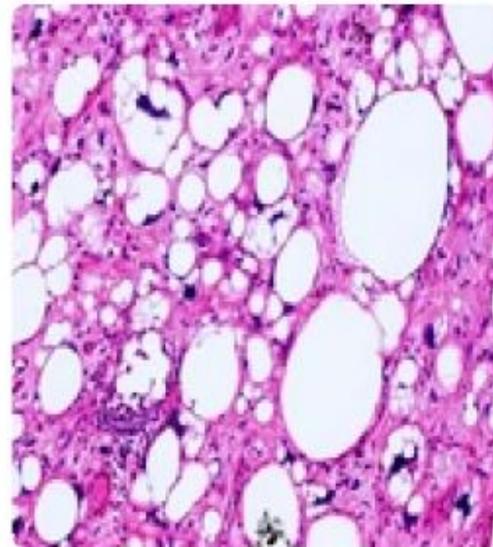
Medical Imaging and AI



Chest X-Ray

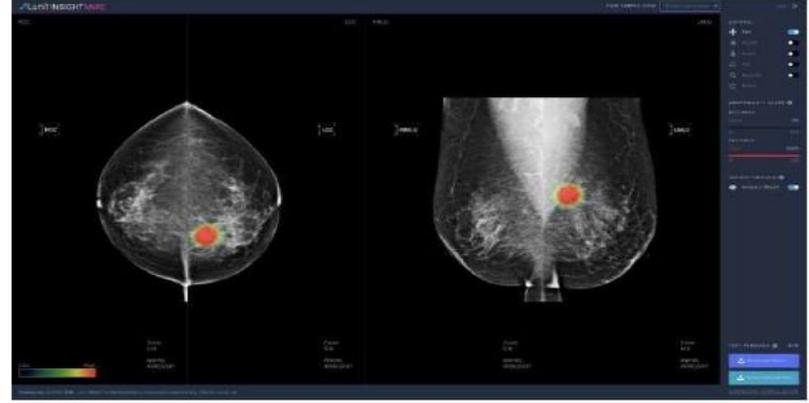


Mammogram



Pathology

Lunit



- 의료 인공지능 스타트업
- 의료 영상을 통한 진단과 치료를 돕는 인공지능 SW 연구 및 개발
- 대표 제품
 - 흉부 X-RAY 판독보조 Lunit INSIGHT CXR
 - 유방촬영 영상 판독보조 Lunit INSIGHT MMG

Background - Medical AI



Lunit Inc.
% John J. Smith, M.D., J.D.
Partner
Hogan Lovells US LLP
555 Thirteenth Street NW
WASHINGTON DC 20004

November 17, 2021

Re: K211678

Trade/Device Name: **Lunit INSIGHT MMG**

Regulation Number: 21 CFR 892.2090

Regulation Name: Radiological computer assisted detection and diagnosis software

Regulatory Class: Class II

Product Code: QDQ

Dated: October 7, 2021

Received: October 7, 2021

Background - Medical AI



Lunit Inc.
% Colin Jacob
15th Floor, 27 Teheran-ro 2-gil
Gangnam-gu, Seoul
REPUBLIC OF KOREA

Re: K211733

November 10, 2021

Trade/Device Name: **Lunit INSIGHT CXR Triage**

Regulation Number: 21 CFR 892.2080

Regulation Name: Radiological computer aided triage and notification software

Regulatory Class: Class II

Product Code: QFM

Dated: October 1, 2021

Received: October 1, 2021

Main - Biostatistical application in clinical study

Overview of clinical study for SaMD

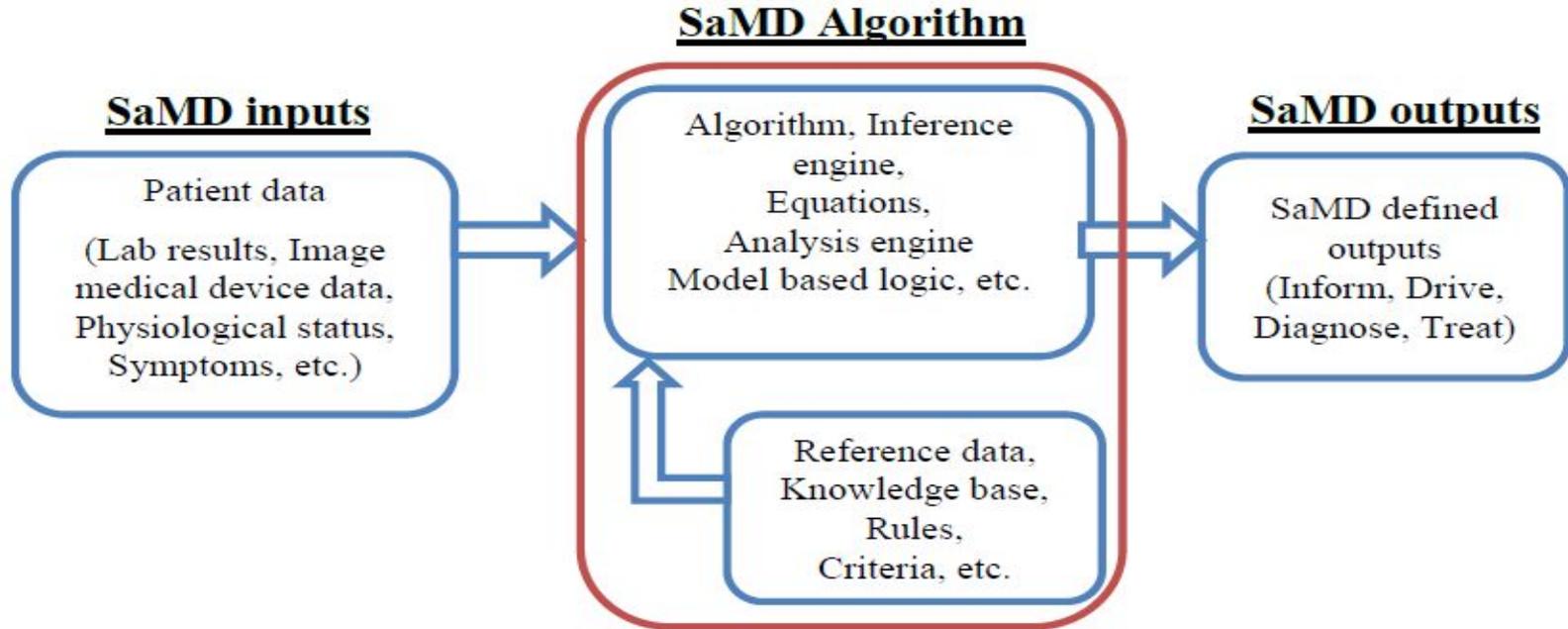
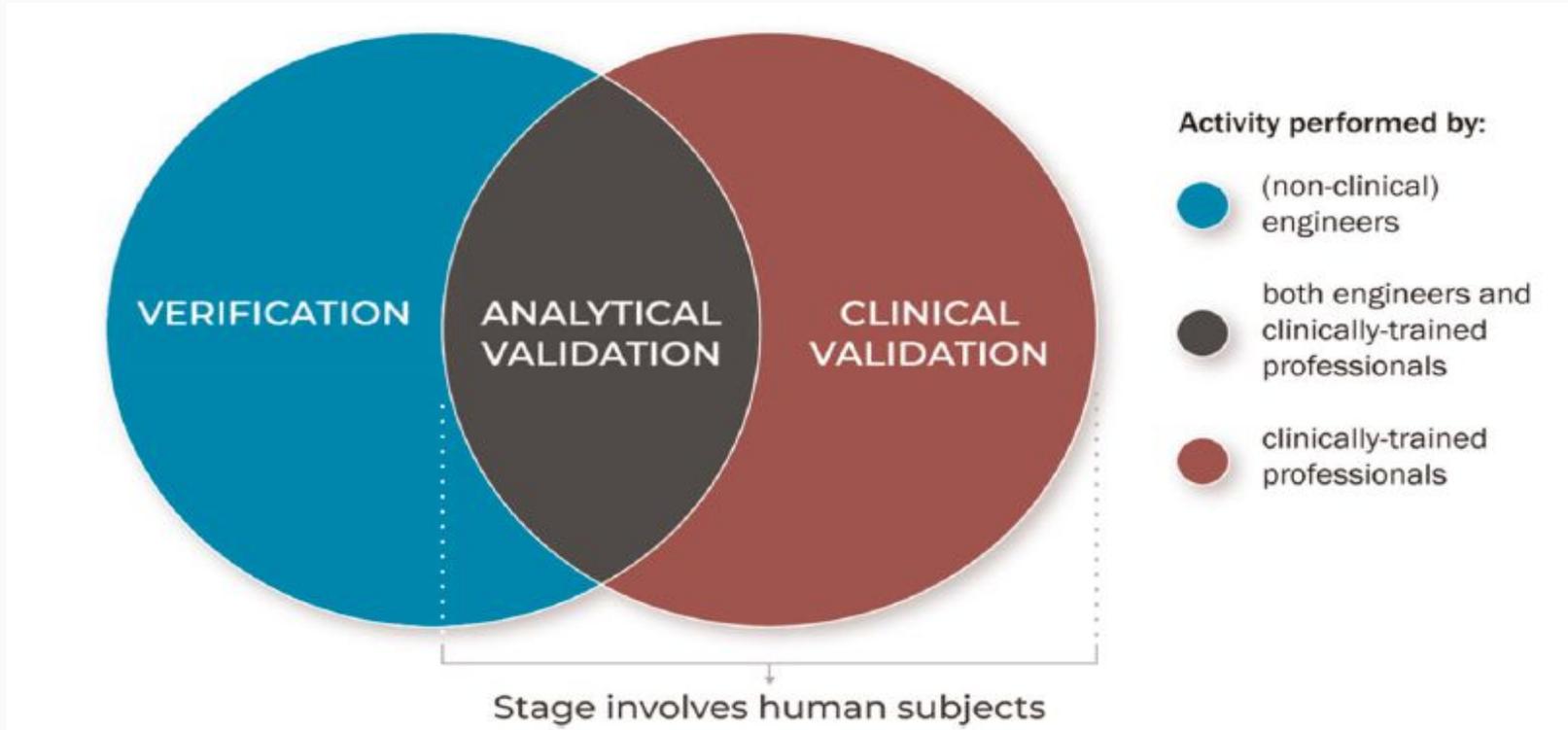


Figure 9 - SaMD Basic Programming Model

Main - Biostatistical application in clinical study

Overview of clinical study for SaMD

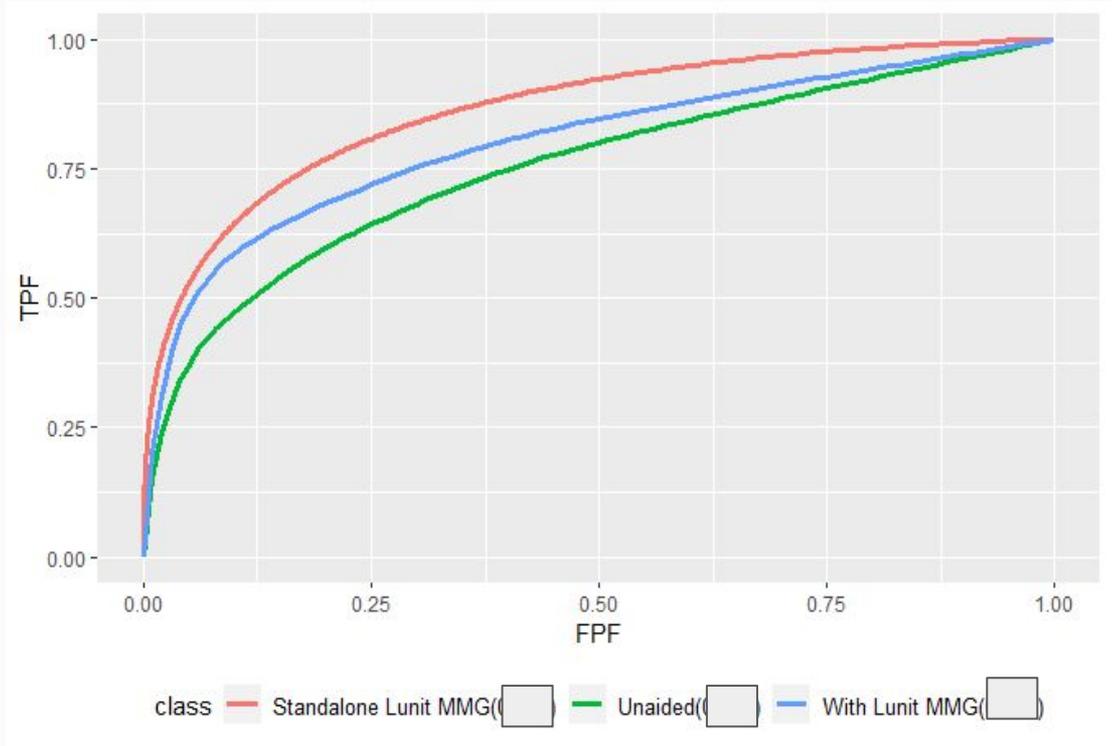


Main - Biostatistical application in clinical study

Statistical inference

Unaided vs With AI

- Performance Improvement with use of AI?



Main - Performance Evaluation

Historical Background (ROC Curve)

- World War II
- Classification of radar signal
 - Noise vs War Plane?

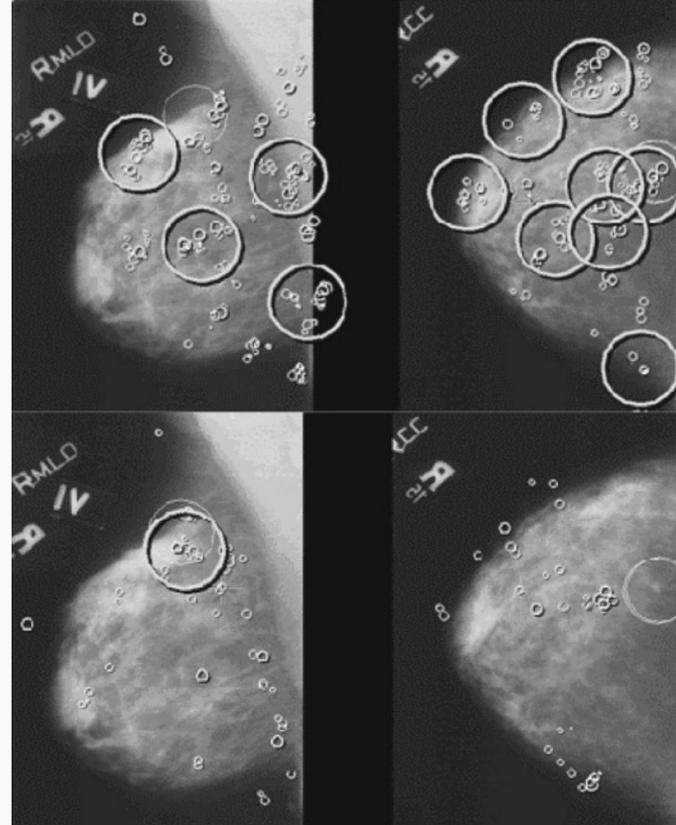


Main - Performance Evaluation

Application in Radiology (ROC Curve)

- Medical Image in Radiology
 - Chest X-ray, MMG, etc...

- Suspicious region on the image
 - Lesion vs Noise?

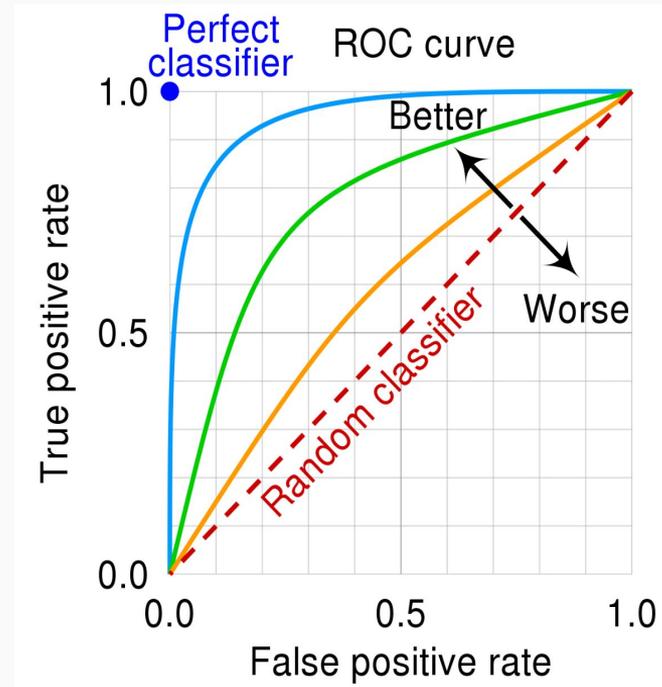


Main - Performance Evaluation

Construction of ROC Curve

- Sensitivity(y - axis)
- 1 - Specificity (x-axis)

		Disease:		
		Sick	Healthy	
Test result:	Positive	True positive (TP)	False positive (FP)	→ PPV
	Negative	False negative (FN)	True negative (TN)	→ NPV
		↓ Sensitivity	↓ Specificity	



Main - Statistical Analysis for Regulatory Approval

Primary endpoint vs Secondary endpoint

Primary Endpoint

- outcome(s) that establish the effectiveness of the device in order to support regulatory action*
- **ROC summary performance metric** is recommended**

Secondary Endpoint

- To demonstrate additional effects after success on the primary endpoint*
- Sensitivity and Specificity is recommended when ROC summary performance metric is used**

Main - Statistical Analysis for Regulatory Approval

Sample size estimation

$H_0 : AUC_1 = AUC_2$ versus $H_1 : AUC_1 \neq AUC_2$

$$n = \frac{\left[Z_{\frac{\alpha}{2}} \sqrt{V_{H0}(\widehat{AUC}_1 - \widehat{AUC}_2)} + Z_{\beta} \sqrt{V_{H1}(\widehat{AUC}_1 - \widehat{AUC}_2)} \right]^2}{[AUC_1 - AUC_2]^2}$$

Factors affecting on sample size

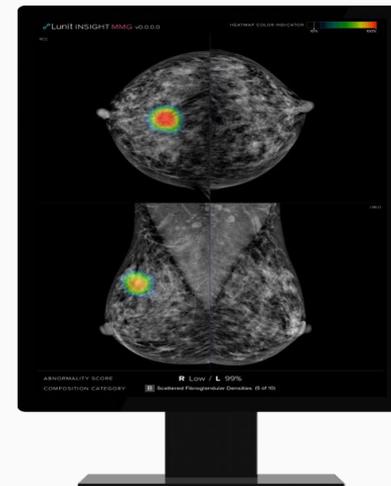
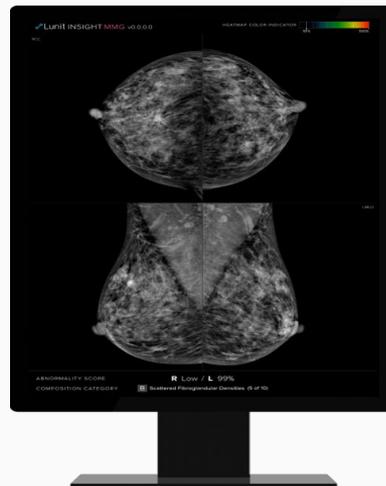
- Predetermined value from previous studies
- Success criteria for the clinical study
- Type I error and Type II error for statistical judgement

Main - Statistical Analysis for Regulatory Approval

Multi-Reader Multi-Case Study

Test 1(Without AI CAD) AUC vs Test 2 (With AI CAD) AUC

$$Y_{ijk} = \mu + \tau_i + R_j + C_k + (\tau R)_{ij} + (\tau C)_{ik} + (RC)_{jk} + (\tau RC)_{ijk} + \varepsilon_{ijk}$$



Main - Statistical Analysis for Regulatory Approval

Standalone Study

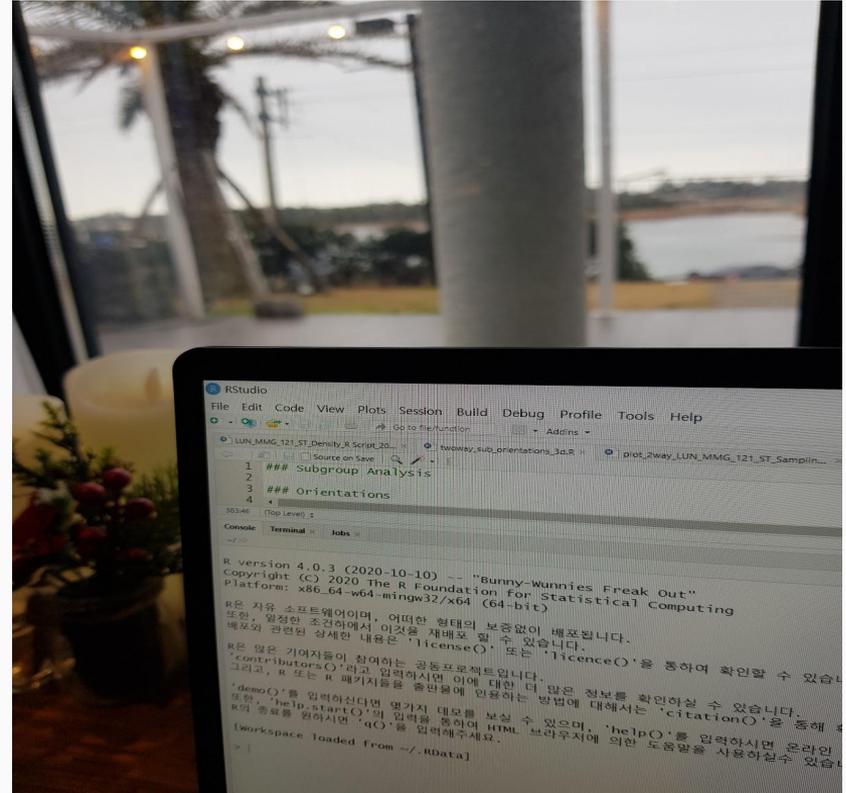
Standalone Performance Assessment

- The performance of **the device by itself**
 - in the absence of any interaction with a clinician
- Estimates how well
 - the device marks **regions of known abnormalities** and
 - the device **avoids marking regions** other than the abnormalities(e.g., normal organ and structures)

Main - Statistical Analysis for Regulatory Approval

Statistical Analysis using R

- Specific R Packages
 - RJaFROC
 - MRMCaov



Main - Statistical Analysis for Regulatory Approval

Protocol

[Hypothesis in the pivotal reader study of the Lunit

- Null hypothesis (H_0): $AUC_{\text{test1}} = AUC_{\text{test2}}$

The average ROC AUC is equivalent when the reading panelist evaluates the image with or without CAD assistance.

- Alternative hypothesis (H_1): $AUC_{\text{test1}} \neq AUC_{\text{test2}}$

The average ROC AUC is not equivalent when the reading panelist evaluates the image with or without CAD assistance.

Main - Statistical Analysis for Regulatory Approval

Protocol

For effect size set as with alpha 0.05 and power 0.9, the number of readers and cases required for ROC AUC comparison are estimated as below. ↵

Number of Reading Panelist ↵	Number of Cases ↵
10 ↵	1015 ↵
11 ↵	640 ↵
12 ↵	468 ↵
13 ↵	369 ↵
14 ↵	305 ↵
15 ↵	260 ↵
16 ↵	226 ↵

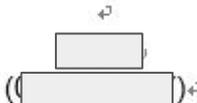
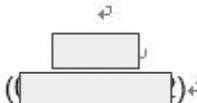
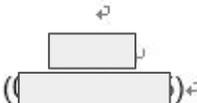
Main - Statistical Analysis for Regulatory Approval

Clinical Study Report

3. Endpoint Analysis

3.1. Primary Endpoint Analysis

[Table 3.1] Patient-level LOS area under the receiver operator characteristic(ROC) curve

LOS	Test1	Test2	Test2-Test1	P-value ²⁾
ROC AUC (95%CI) ¹⁾				0.0001

LOS: Level of Suspicion, ROC AUC: Area under the receiver operating characteristic (ROC) curve

CI: Confidence Interval

Test1: Conventional reading without CAD assistance

Test2: Second-reading with CAD assistance

- 1) Readers and images were treated as random effects. Non-parametric trapezoidal method was used to estimate the area under the ROC curve (AUC)
- 2) P-value for statistical difference of ROC AUC (Test2 – Test1)

Main - Statistical Analysis for Regulatory Approval

510(k) Summary

Lunit INSIGHT MMG (K211678)

This 510(k) summary of safety and effectiveness information is prepared in accordance with the requirements of 21 CFR §807.92.

Date Prepared: November 15, 2021

I. SUBMITTER

Manufacturer:

Lunit Inc.

15 Floor, 27 Teheran-ro 2-gil, Gangnam-gu, Seoul, Republic of Korea 06241

Phone +82-2-2138-0827

Fax +82-2-2135-5413

Main - Statistical Analysis for Regulatory Approval

The primary endpoint of the standalone performance measured the Lunit INSIGHT MMG performance compared to radiologist performance alone. ROC AUC in the standalone performance analysis was 0.903 (95% CI: 0.889-0.917) with statistical significance ($p < 0.0001$), which demonstrates improvement compared to the interpretation performance of radiologists when reading mammograms unaided. The secondary exploratory endpoints are two types of LROC AUC, sensitivity and specificity and these results are as follow:

[Standalone Analysis Results]

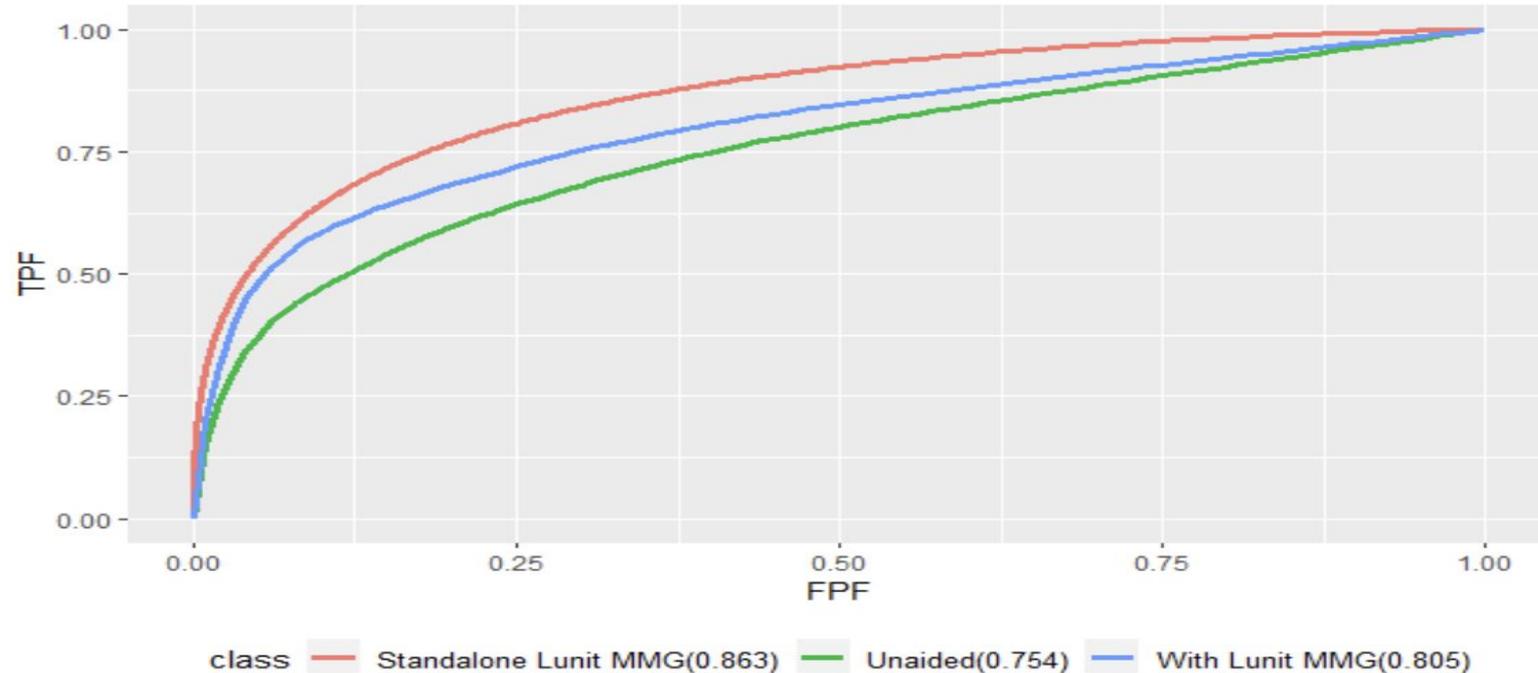
ROC AUC [95% CI]	0.903 [0.889, 0.917] * $p < 0.0001$
Type I LROC AUC [95% CI]	0.781 [0.751, 0.812]
Type II LROC AUC [95% CI]	0.792 [0.763, 0.822]
Sensitivity (%) [95% CI]	85.74 [82.95, 88.53]
Specificity (%) [95% CI]	75.62 [73.64, 77.60]

Clinical Testing – Reader Study

Clinical Performance Assessment was conducted to evaluate effectiveness of Lunit INSIGHT MMG in the assistance of detection and diagnosis of breast cancer during screening mammography interpretation. A retrospective, multi-reader multi-case (MRMC) study was conducted comparing the reading panel's interpretation performance with and without the assistance of the Lunit INSIGHT MMG during the screening mammography interpretation.

Main - Statistical Analysis for Regulatory Approval

[Plot Comparison of the standalone performance, CAD unaided and aided reader interpretation]



Main - Use of AI in Work (typeset.io)



Overview

Full-text

Citations (21.9K)

Related Papers (5)

attention over the output of the encoder stack. Similar to the encoder, we employ residual connections around each of the sub-layers. We also modify the sub-layer in the decoder stack to prevent positions from attending to subsequent positions. This masking, combined with fact that the output embeddings are offset by one position, ensures that the predictions for position i can depend only on the known outputs at positions less than i .

 Explain math & table

 116% 

3.2 Attention

An attention function can be described as mapping a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key.

3.2.1 Scaled Dot-Product Attention

We call our particular attention "Scaled Dot-Product Attention" (Figure 2). The input consists of queries and keys of dimension d_k , and values of dimension d_v . We compute the dot products of the

Main - Use of AI in Work (typeset.io)

 Copilot

 English (en) ▾

Explain the text

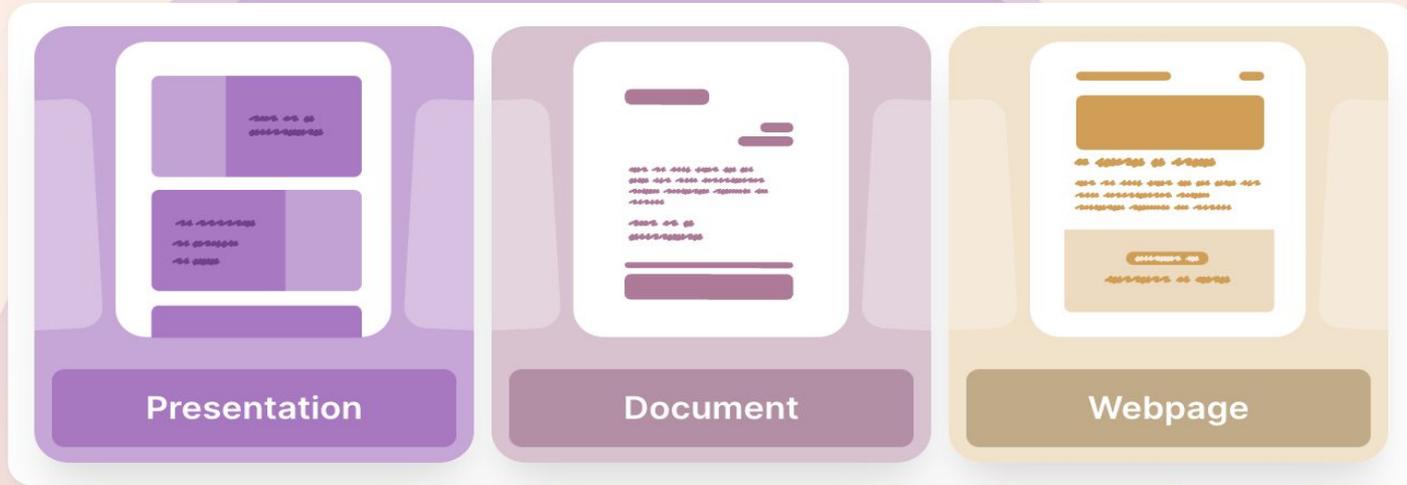
attention function can be described as mapping a query and a set of key-value pairs to an output,

7:43 pm

The highlighted text refers to an attention mechanism used in the proposed network architecture. The attention function is a mathematical operation that takes a query and a set of key-value pairs as inputs and produces an output. The query is a vector that represents the current state of the decoder, while the key-value pairs are vectors that represent the states of the encoder. The attention function computes a similarity score between the query and each key vector, using a dot product or a more complex function. The scores are then normalized using a softmax function to obtain a probability distribution over the key vectors. The values are weighted by the probabilities and summed up to produce the output, which is a context vector that captures the most relevant information from the encoder for the current decoding step. The attention mechanism allows the decoder to focus on different

Main - Use of AI in Work (gamma.app)

👋 Hi SooYoung, I'm your AI design partner. What would you like to create today?



Main - Use of AI in Work (gamma.app)

Contents

- Introduction to AFROC in Medical AI
- The Basics of AFROC
- Calculating AFROC
- The significance of FOM in Medical AI
- Application of AFROC across Medical AI Systems
- Potential Future Work in AFROC
- Conclusion

Introduction to AFROC in Medical AI

Theme Share Present 0

Introduction to AFROC in Medical AI

AFROC, or Alternative Free-Response Receiver Operating Characteristic, is a statistical method used to evaluate and validate medical AI systems. In this presentation, you'll learn how AFROC FOM is used in the Medical AI field and how to estimate it.

by **SooYoung Kwak**
Last edited about 1 month ago



Main - Use of AI in Work (gamma.app)

Contents

- Introduction to AFROC in Medical AI
- The Basics of AFROC
- Calculating AFROC
- The significance of FOM in Medical AI
- Application of AFROC across Medical AI Systems
- Potential Future Work in AFROC
- Conclusion

Introduction to AFROC in Medical AI

Theme

Share

Present

0

...

The Basics of AFROC



AFROC Definition

AFROC is a statistical method used to evaluate diagnostic systems in medicine. It measures the performance of the system when the output is free-form and the response is not binary.



Use Cases of AFROC FOM

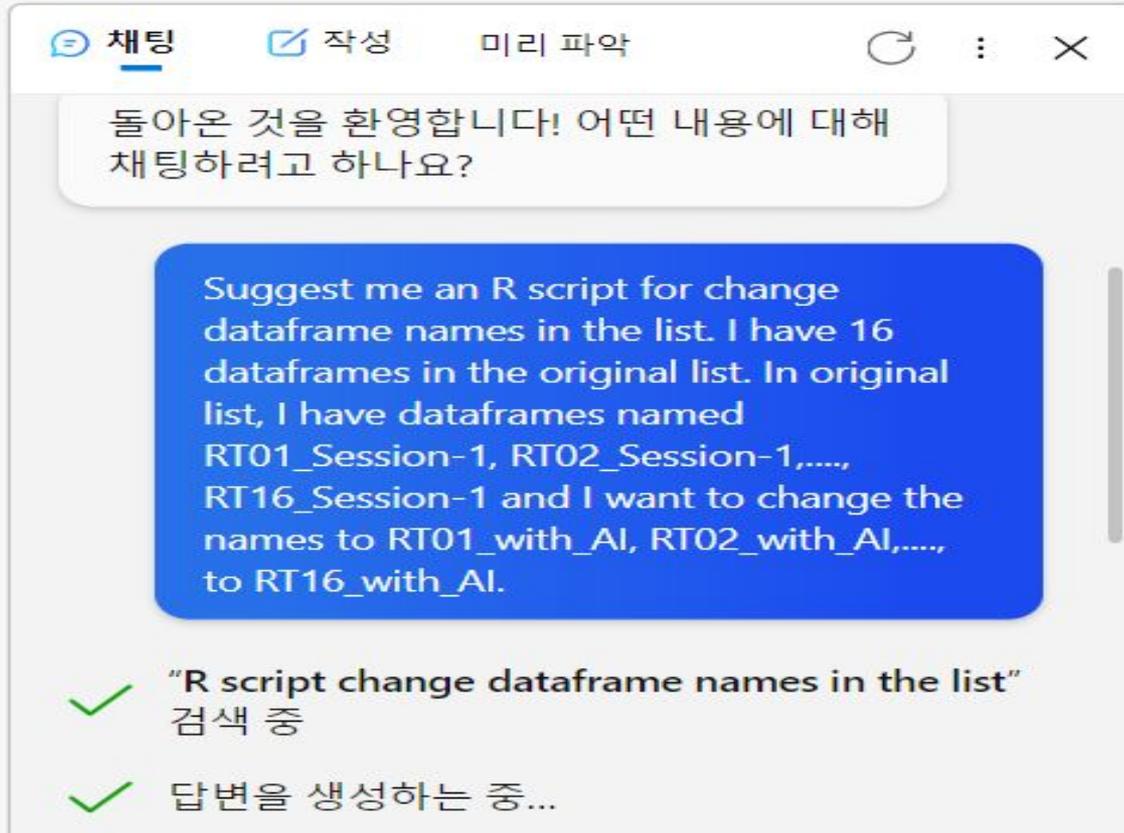
AFROC FOM is used in Medical AI Field to measure the performance of diagnostic systems. The method can evaluate the performance of systems designed to predict outcomes such as readmission or mortality risk.



The Role of AFROC FOM in Medical Research

AFROC FOM allows researchers to better understand the performance of the models being developed, which then enables them to focus on improving models and identify areas that need improvement.

Main - Use of AI in Work (Bing chatGPT)



The screenshot shows a chat window with a header bar containing icons for chat, compose, and search, along with window controls. A system message in Korean asks for a topic to chat about. The user's query, in English, asks for an R script to rename 16 dataframes in a list. Below the query, two green checkmarks indicate the search status: "R script change dataframe names in the list" is being searched, and the response is being generated.

채팅 작성 미리 파악 ↻ ⋮ ✕

돌아온 것을 환영합니다! 어떤 내용에 대해 채팅하려고 하나요?

Suggest me an R script for change dataframe names in the list. I have 16 dataframes in the original list. In original list, I have dataframes named RT01_Session-1, RT02_Session-1,....., RT16_Session-1 and I want to change the names to RT01_with_AI, RT02_with_AI,....., to RT16_with_AI.

✓ "R script change dataframe names in the list"
검색 중

✓ 답변을 생성하는 중...

Main - Use of AI in Work (Bing chatGPT)

✓ 답변을 생성하는 중...

Here is an R script that you can use to change the names of dataframes in a list. You can use the `lapply()` function to loop through each dataframe in the list and rename it using the `setNames()` function. Here's how you can do it:

```
# Create a list of dataframes
original_list <-
  list(RT01_Session-1,
        RT02_Session-1, RT03_Session-1,
        RT04_Session-1, RT05_Session-1,
        RT06_Session-1, RT07_Session-1,
        RT08_Session-1, RT09_Session-1,
        RT10_Session-1, RT11_Session-1,
        RT12_Session-1, RT13_Session-1,
        RT14_Session-1, RT15_Session-1,
        RT16_Session-1)

# Define new names for dataframes
new_names <- c("RT01_with_AI",
               "RT02_with_AI", "RT03_with_AI",
               "RT04_with_AI", "RT05_with_AI",
               "RT06_with_AI", "RT07_with_AI",
               "RT08_with_AI", "RT09_with_AI",
               "RT10_with_AI", "RT11_with_AI",
               "RT12_with_AI", "RT13_with_AI",
               "RT14_with_AI", "RT15_with_AI",
               "RT16_with_AI")
```

Main - Use of AI in Work (Bing chatGPT)

Can you suggest an regex to do this task?

- ✓ "regex to rename dataframe names in R" 검색 중
- ✓ 답변을 생성하는 중...

Here is a regex that you can use to rename the dataframe names in R:

```
# Create a list of dataframes
original_list <-
list(RT01_Session-1,
RT02_Session-1, RT03_Session-1,
RT04_Session-1, RT05_Session-1,
RT06_Session-1, RT07_Session-1,
RT08_Session-1, RT09_Session-1,
RT10_Session-1, RT11_Session-1,
RT12_Session-1, RT13_Session-1,
RT14_Session-1, RT15_Session-1,
RT16_Session-1)

# Define new names for dataframes
using regex
new_names <- gsub(₩"Session₩",
₩"with_AI₩",
names(original_list))

# Use lapply() function to loop
through each dataframe in the
```

Main - Use of AI in Work (Bing chatGPT)

```
# Define new names for dataframes
using regex
new_names <- gsub(₩"Session₩",
₩"with_AI₩",
names(original_list))

# Use lapply() function to loop
through each dataframe in the
list and rename it using
setNames() function
new_list <-
lapply(seq_along(original_list),
function(i)
setNames(original_list[[i]],
new_names[i]))

# Print the new list of
dataframes
new_list
```

This script will create a new list of dataframes with the new names that you specified using regex. The `gsub()` function is used to replace the word "Session" with "with_AI" in the original dataframe names. I hope this helps!

자세한 정보:

2 / 20 ●

1. stackoverflow.com

2. stackoverflow.com

3. stackoverflow.com

— 간략히 보기

Main - Use of AI in Work (Bing chatGPT)

```
57 ### Create a list of file names
58 files2 <- list.files(path = "C:/Users/USER/Desktop/Stat_Analysis/ ",
59                      pattern = "*.csv", full.names = TRUE)
60
61 list_df2 <- lapply(files2, read.csv)
62
63 # Assign names to the list elements based on file names
64 names(list_df2) <- gsub("-\\d_Session-(\\d).csv", "_\\1", basename(files2))
65
```

Summary

2022 WORK REVIEW

Statistical Analysis Result

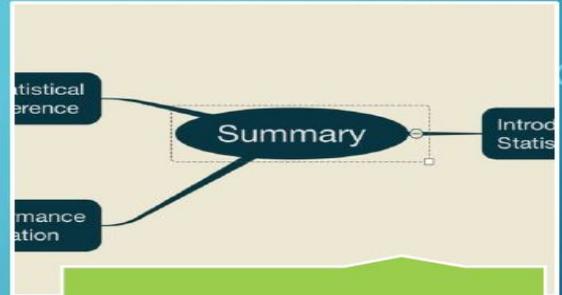
AN INFORMATION	LUN_CCT_411
Multi-Reader, Multi-Case (MRMC), Retrospective, Exploratory	
Examine the Feasibility of the Investigational Device in the Pulmonary Nodule in Chest CT	
	1.0
	1.0
	1.0

Stat Analyses

Interim Change Control Plan for Lunit INSIGHT

Indication
Indications For Use
Device Description
are pre-specifications
Algorithmic changes
1.1.1 Changes in training process
1.1.2 Changes in network architecture
1.1.3 Changes in preprocessing
1.1.4 Changes in the data
Input
1.2.1 Device coverage
1.2.2 Analyze change
Intention

FDA PCCP

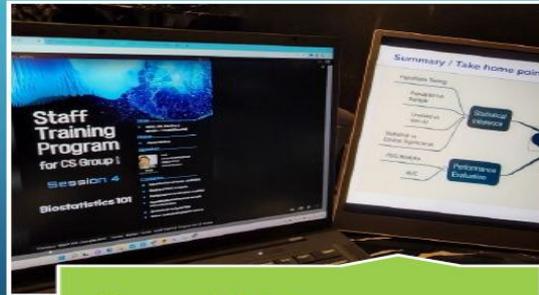


Seminar

Ch14. Deep Learning in Breast Cancer Screening

Information
File:
CH14 DL in Breast Imaging.pdf
Link: (기타문 부분은 이 책에서 Medical Imaging의 Chapter 14)
<https://www.amazon.com/Artificial-Intelligence-Medical-Imaging-Opportunities/dp/3319988776>

Article Review

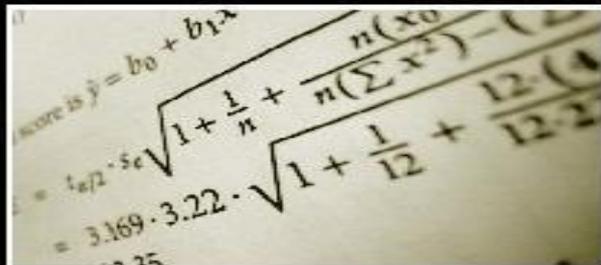


Staff Training



Writing

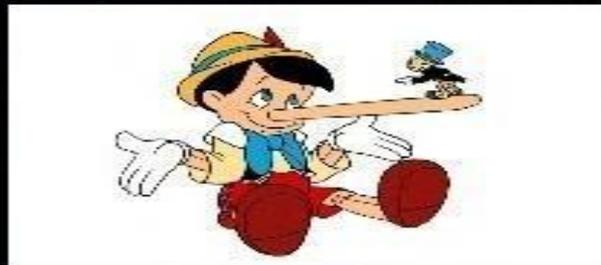
STATISTICIAN



What my friends think I do



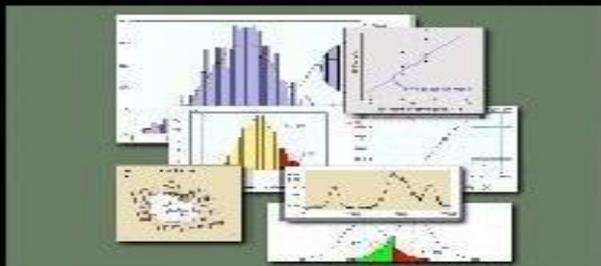
What my parents think I do



What society thinks I do



What my boss thinks I do



What I think I do



What I really do