Journal of
Breast Cancer **JBC**

Study Protocol

[Check for updates]

Artificial Intelligence for Breast Cancer Screening in Mammography (AI-STREAM): A Prospective Multicenter Study Design in Korea Using AI-Based CADe/x

Yun-Woo Chang [1,*], **Jin Kyung An** [2], **Nami Choi** [3], **Kyung Hee Ko** [4], **Ki Hwan Kim** [5], **Kyunghwa Han** [6], **Jung Kyu Ryu** [7,*]

[1]Department of Radiology, Soonchunhyang University Seoul Hospital, Soonchunhyang University College of Medicine, Seoul, Korea
[2]Department of Radiology, Nowon Eulji University Hospital, Eulji University School of Medicine, Seoul, Korea
[3]Department of Radiology, Konkuk University Medical Center, Konkuk University School of Medicine, Seoul, Korea
[4]Department of Radiology, CHA Bundang Medical Center, Seongnam, Korea
[5]Lunit, Seoul, Korea
[6]Department of Radiology, Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea
[7]Department of Radiology, Kyung Hee University Hospital at Gangdong, College of Medicine, Kyung Hee University, Seoul, Korea

Correspondence to
Jung Kyu Ryu
Department of Radiology, Kyung Hee University Hospital at Gangdong, College of Medicine, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Korea.
Email: oddie2@naver.com

Yun-Woo Chang
Department of Radiology, Soonchunhyang University Seoul Hospital, Soonchunhyang University College of Medicine, 59 Daesagwan-ro, Yongsan-gu, Seoul 04401, Korea.
Email: ywchang@schmc.ac.kr

*These authors contributed equally to this work

ORCID iDs
Yun-Woo Chang
https://orcid.org/0000-0001-9704-8112
Jin Kyung An
https://orcid.org/0000-0003-1086-426X
Nami Choi
https://orcid.org/0000-0002-2703-8282

ABSTRACT

Purpose: Artificial intelligence (AI)-based computer-aided detection/diagnosis (CADe/x) has helped improve radiologists' performance and provides results equivalent or superior to those of radiologists' alone. This prospective multicenter cohort study aims to generate real-world evidence on the overall benefits and disadvantages of using AI-based CADe/x for breast cancer detection in a population-based breast cancer screening program comprising Korean women aged ≥ 40 years. The purpose of this report is to compare the diagnostic accuracy of radiologists with and without the use of AI-based CADe/x in mammography readings for breast cancer screening of Korean women with average breast cancer risk.
Methods: Approximately 32,714 participants will be enrolled between February 2021 and December 2022 at 5 study sites in Korea. A radiologist specializing in breast imaging will interpret the mammography readings with or without the use of AI-based CADe/x. If recall is required, further diagnostic workup will be conducted to confirm the cancer detected on screening. The findings will be recorded for all participants regardless of their screening status to identify study participants with breast cancer diagnosis within both 1 year and 2 years of screening. The national cancer registry database will be reviewed in 2026 and 2027, and the results of this study are expected to be published in 2027. In addition, the diagnostic accuracy of general radiologists and radiologists specializing in breast imaging from another hospital with or without the use of AI-based CADe/x will be compared considering mammography readings for breast cancer screening.
Discussion: The Artificial Intelligence for Breast Cancer Screening in Mammography (AI-STREAM) study is a prospective multicenter study that aims to compare the diagnostic

Generated by xmlinkpress

Kyung Hee Ko (iD)
https://orcid.org/0000-0003-1373-2502
Ki Hwan Kim (iD)
https://orcid.org/0000-0001-7684-235X
Kyunghwa Han (iD)
https://orcid.org/0000-0002-5687-7237
Jung Kyu Ryu (iD)
https://orcid.org/0000-0001-8195-0785

accuracy of radiologists with and without the use of AI-based CADe/x in mammography readings for breast cancer screening of women with average breast cancer risk. AI-STREAM is currently in the patient enrollment phase.

Trial Registration: ClinicalTrials.gov Identifier: NCT05024591

Keywords: Artificial Intelligence; Breast; Clinical Trial; Digital Mammography; Early Detection of Cancer

INTRODUCTION

Breast cancer in women has a particularly high incidence globally and ranks first among the cancers in Korean women [1,2]. Breast cancer is prevalent in women aged 35–64 years (accounting for 29.1% of all female cancer patients), which is the most economically active age group [2]. Early detection of breast cancer improves patient prognosis and survival [3]. Therefore, in Korea, women aged ≥ 40 years are invited to undergo breast cancer screening with mammography every 2 years as part of the national cancer screening program [4].

However, there have been several challenges with mammography-based breast cancer screening, including false positives and negatives. First, patients who were not identified through breast cancer screening have been reported to have missed cancer diagnosis within 1 year after screening. Second, several false-positive recalls are also reported; of approximately 530,000 women who underwent additional tests after screening, only 1.3% were confirmed to have breast cancer within 1 year [5]. In particular, detection of breast cancer is more difficult with higher breast density than with lower breast density owing to the low sensitivity of mammography with a high-density breast, making it difficult to differentiate between a dense tissue and a breast lesion [6]. Notably, 52.6% of Korean women and 65.5% of women in their 40s and 50s have dense breasts [7]. Therefore, achieving a high breast cancer detection rate (CDR) in cases of dense breasts is necessary to maximize the effectiveness of breast cancer screening. There lies an inter-reader difference in the performance of mammography [8]. According to a study that evaluated mammogram reading performance for breast cancer screening in the United States, approximately 10% of all physicians had a CDR below the acceptable range and more than 25% of physicians had a recall rate exceeding the acceptable range [9]. These findings indicate the difficulty in mammogram reading and support the need for artificial intelligence (AI)-based computer-aided detection/diagnosis (CADe/x) to help radiologists establish an accurate diagnosis.

Computer-aided detection (CAD) was first developed for educational and academic purposes, and it was commercialized in 1998 with the US Food and Drug Administration approving the ImageChecker system, the first commercial CAD system for mammography [10]. CAD differentiates structures and sections on medical images based on complex pattern recognition, but its low specificity and undesirably high recall rate limit its use in clinical practice [10]. Subsequent advancements in AI technology have improved CAD accuracy. In particular, CAD using the convolution neural network, a deep learning technology optimized for image analysis and classification, provides diagnostic performance that is equivalent or superior to that of radiologists' [11]. AI-based CADe/x could help in the following situations: breast cancer not identified through screening, excessive recalls for further testing, low sensitivity in cases of dense breasts, and inter-reader variability. Lunit INSIGHT MMG (Lunit, Seoul, Korea), a

software that aids mammogram reading by detecting breast cancer, was developed by training an AI algorithm using breast cancer mammography data that included more than 200,000 cases analyzed in Korea, the United States, and the United Kingdom [12]. It has received authorization for use from the Korean Ministry of Food and Drug Safety (MFDS).

Artificial Intelligence for Breast Cancer Screening in Mammography (AI-STREAM) is designed to prospectively study a multicenter cohort and generate real-world evidence on the overall benefits and disadvantages of AI-based CADe/x using Lunit INSIGHT MMG AI-based CADe/x for breast cancer detection in a population-based breast cancer screening program comprising Korean women aged ≥ 40 years. The purpose of this study is to compare the diagnostic accuracy of radiologists with and without the use of AI-based CADe/x in mammography readings for breast cancer screening of Korean women with average breast cancer risk.

METHODS

The study is financially supported by a grant from the Korea Health Industry Development Institute with its third Korea Medical Device Development Fund in 2020 (No. KMDF_PR_20200910_0300). This study protocol was approved by the Institutional Review Boards (IRBs) (First approval: January 5, 2021) of all participating centers (Kyung Hee University Hospital at Gangdong, Soonchunhyang University Seoul Hospital, CHA Bundang Medical Center, Konkuk University Medical Center, and Nowon Eulji Medical Center), and written consent for data publication was and will be obtained from all participants.

Study design

AI-STREAM is a prospective multicenter cohort study. Women eligible for national cancer screening in the relevant year who read the participant recruitment brochure and read and signed the Participant Information Sheet and Informed Consent Form will be recruited in this study. Participants will be recruited from 5 academic hospitals in South Korea: Kyung Hee University Hospital at Gangdong, Soonchunhyang University Seoul Hospital, CHA Bundang Medical Center, Konkuk University Medical Center, and Nowon Eulji Medical Center. Approximately 32,714 study participants will be enrolled from February 2021 through December 2022. An overview of the study design is shown in **Figure 1**, and an overview of the study timeline is presented in **Figure 2**.

Since cancer registry data are only available at least 26 months after the year of interest, databases will be reviewed twice in 2026 and 2027 for calculating the diagnostic accuracy endpoints, which are based on breast cancer diagnosis within 1 year and 2 years of screening. An interim analysis is planned in 2026 after reviewing the registry databases for the first time for breast cancer diagnosis. Final analyses will be performed after all participant data have been collected and cleaned, and database lock has occurred. Data will be collected using an electronic case report form (eCRF).

Ethical consideration

The IRB of the 5 participating institutions approved this study. This protocol was registered in the International Clinical Trial Registry Platform (Clinical Research Information Service, ClinicalTrials.gov Identifier: NCT05024591) in August 2021. The first participant was enrolled in February 2021. At the time of manuscript submission (December 2021), active inclusion
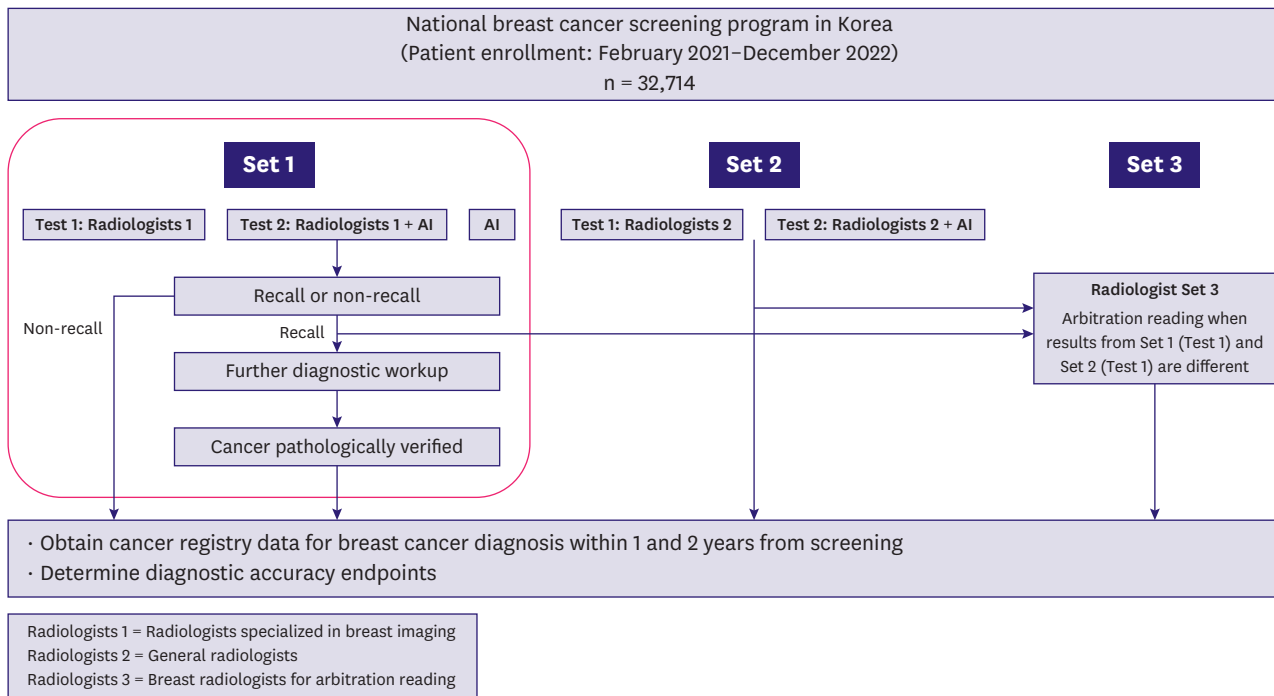
National breast cancer screening program in Korea
(Patient enrollment: February 2021–December 2022)
n = 32,714

Set 1

Test 1: Radiologists 1 | Test 2: Radiologists 1 + AI | AI

Non-recall

Recall or non-recall

Recall

Further diagnostic workup

Cancer pathologically verified

Set 2

Test 1: Radiologists 2 | Test 2: Radiologists 2 + AI

Set 3

Radiologist Set 3
Arbitration reading when results from Set 1 (Test 1) and Set 2 (Test 1) are different

· Obtain cancer registry data for breast cancer diagnosis within 1 and 2 years from screening
· Determine diagnostic accuracy endpoints

Radiologists 1 = Radiologists specialized in breast imaging
Radiologists 2 = General radiologists
Radiologists 3 = Breast radiologists for arbitration reading

Figure 1. Overview of the study design.
AI = artificial intelligence.

Data collected after screening mammography:
· Participant-reported demographics and medical history
· (Set 1) Radiologist-collected mammographic variables
· Other mammographic and pathological features from medical records

Data collected retrospectively:
· (Set 2 and Set 3) Radiologist-collected mammographic variables

First registry data collection:
· Breast cancer diagnosis within 1 year from screening
· Survival status

Second registry data collection:
· Breast cancer diagnosis within 2 years from screening
· Survival status

Feb 1, 2021 | Recruitment period | Dec 31, 2022 | Dec 2023 | Dec 2024 | 2026 | 2027

Interim analysis and report

Final analysis and report

Example scenarios

Scenario A — Within 6 months
Scenario B — Within 12 months
Scenario C — Within 24 months
Scenario D — Follow-up for 24 months
Scenario E
Scenario F

✖ Breast cancer screening (study enrollment)
✖ Cancer diagnosis
● Death
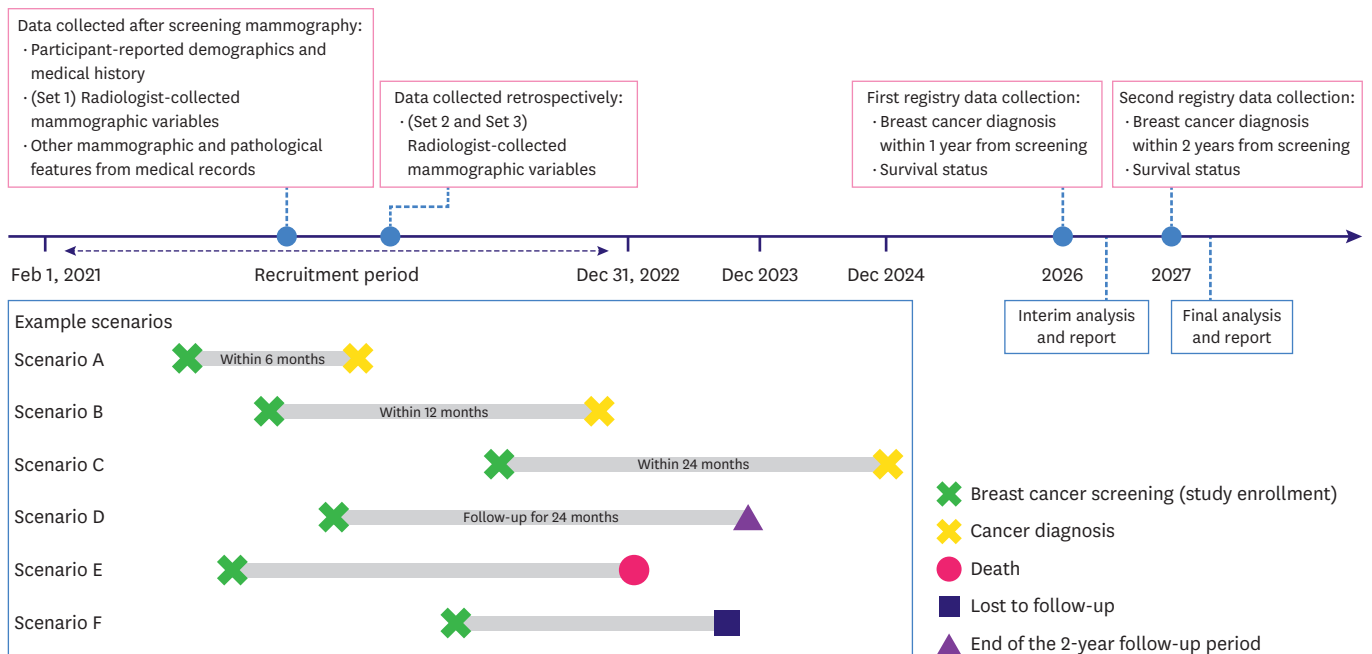■ Lost to follow-up
▲ End of the 2-year follow-up period

Figure 2. Overview of the study timeline.
Scenarios A and B: Enrolled and diagnosed with cancer at 6 or 12 months; classified as having breast cancer diagnosis within 1 year from screening. Scenario C: Diagnosed with cancer within 24 months of enrollment; classified as having breast cancer diagnosis within 2 years from screening. Scenario D: Not diagnosed with cancer until 24 months after enrollment. The results of Scenarios A–D can be termed positive or negative depending on the cancer diagnosis. Scenario E: Died during the study because of any reason and cannot be included in the study. Scenario F: Lost to follow-up and cannot be included in the study because of no available records.

of participants will be ongoing in 5 centers, and a total of 7,000 women would have been enrolled and have undergone the first round of screening. Approximately 32,714 study participants will be enrolled from February 2021 through December 2022, and the study results are expected to be presented in 2027.

Protection of human participants

The study will be conducted in accordance with the International Council for Harmonization Good Clinical Practice (as applied to observational research), all applicable participant privacy requirements, and the ethical principles outlined in the Declaration of Helsinki 2013, including but not limited to the following: a) Central IRB and local IRB review and approval of study protocol and any subsequent amendments and b) investigator reporting requirements. As no additional study-specific tests were required for participation in this study, no further physical risk affected participants. The investigational data that will be collected include mammograms and their readings, medical records (pathological type and stage of breast cancer), and a questionnaire. Such information is used only for evaluating breast cancer diagnostic accuracy with breast imaging and breast cancer-related studies, and the collected information is thoroughly controlled according to the Personal Data Protection Act and cannot be disclosed externally for purposes other than the study. There will be no secondary use of the data collected outside this study. The personal information collected will be used for research purposes and will not be used for any purpose other than this study. All imaging data and readings will be recorded and saved as de-identified data by deleting personally identifiable information such as the name and enrollment number. In addition, personal records (name, date of birth, resident registration number, participant enrollment number), address information, and contact information were collected. However, to avoid the possible disclosure of personal information, complete confidentiality will be ensured while using medical records and imaging data of participants based on the confidentiality section. In this study, mammograms will be additionally analyzed with the AI-based CADe/x, and the findings will be used as reference for a radiologist during readings, along with the original mammograms. Therefore, any harm or damage to a participant in this study will not be possible.

Study population

Women aged ≥ 40 years eligible for national cancer screening in the relevant year who read the study participant recruitment brochure and read and signed the Participant Information Sheet and Informed Consent Form will be recruited in this study. Women must meet all the following inclusion criteria to be enrolled in the study: eligible for national cancer screening in the relevant year and visited the site for breast cancer screening, provided consent for study participation using the informed consent form, and completed a study questionnaire. Participants who met any of the following criteria were excluded from the study: history of or current breast cancer, current pregnancy or planned pregnancy over the next 1 year, history of breast surgery (mammoplasty or insertion of a foreign substance such as paraffin or silicon), and underwent mammography for diagnostic purposes (**Supplementary Data 1**).

Image acquisition

Mammography in 2 standard image planes is performed using digital mammography units. Mammograms of participants will be exported to the study platform (BEST image) to enter the reading results after de-identification. The radiologist will mark the collected variables using the BEST image platform after reading the mammograms with and without the use of AI-based CADe/x.

The AI-based CADe/x that will be used is the commercially released Lunit INSIGHT MMG, which detects areas suggestive of breast cancer on mammography, marks areas suggestive of malignant lesions, and displays the probability of the presence of malignant lesions to aid the interpreting physician's diagnosis. The Lunit INSIGHT MMG was developed by training an AI algorithm using breast cancer mammography data (> 50,000 cases), and the total training data included more than 200,000 cases analyzed in Korea, the United States, and the United Kingdom [12]. A clinical study on Lunit INSIGHT MMG was conducted to obtain authorization for use from the Korean MFDS. Mammography reading with Lunit INSIGHT MMG significantly improved the accuracy of breast cancer detection by up to 12.6% and reduced the need for additional tests, with a 13.8% reduction in the non-malignant recall rate [12].

Furthermore, the accuracy of the Lunit INSIGHT MMG in detecting a malignant lesion was 96% area under the receiver operating characteristics curve (AUROC), which was superior to that of other products [13].

Information on the following variables will be collected: breast density (according to Breast Imaging-Reporting and Data System [BI-RADS] 5th Edition [14]); recall (no/yes); recall location (no/left/right/both); and malignant scale (definite normal/benign/probably benign (> 0% but ≤ 2%), low suspicion of malignancy (> 2% to ≤ 10%), moderate suspicion of malignancy (> 10% to ≤ 50%), high suspicion of malignancy (> 50% to < 95%), and highly suggestive of malignancy (≥ 95%). Stand-alone AI-based CADe/x results will also be recorded as positive or negative based on the abnormality scores (**Supplementary Data 1**).

Image interpretation

In Korea, a single radiologist performed mammography readings. As part of the standard screening procedure (Set 1), mammograms will be read and findings will be recorded by radiologists specializing in breast imaging without the use of AI-based CADe/x. Radiologists in this study are experts in breast imaging at an academic hospital for more than 10 years after fellowship training. Mammograms will be interpreted by these radiologists without the use of AI-based CADe/x, and findings will be recorded in the study platform (BEST image) (Test 1). After completion of Test 1, the radiologist will read the mammogram with the use of AI-based CADe/x, and the findings will be recorded based on the radiologist's decision after considering both with and without AI-based CADe/x findings (Test 2). After Test 1 is completed, the reader will proceed to Test 2 and cannot return to Test 1 and re-read after the CADe/x result is reviewed. After completion of Test 2 reading, results cannot be corrected. Results from Test 2 will help determine the requirement of further diagnostic workup. If recall is required (per usual care), further diagnostic workup will be conducted to confirm the cancer detected at screening.

Stand-alone AI-based CADe/x data will also be collected. After completing the standard screening procedure in Set 1, using several situational groups (Set 2 and Set 3) for comparison, the diagnostic accuracy will be performed independently and retrospectively. The results from Set 2 and Set 3 will not impact the clinical decisions associated with the care of the study participants. In Set 2, mammograms from the same participants will be read by general radiologists with and without the use of AI-based CADe/x, and the findings will be recorded in the study platform (BEST image). General radiologists are radiologists not specializing in breast imaging. The participant and reading process are the same as in Set 1. Both radiologist groups do not have prior experience using the CAD program in their practice.

In Set 3, another radiologist specializing in breast imaging from a different hospital will participate in arbitration reading for discordant cases. An arbitration reading will be conducted for discordant cases in which the reading results of the breast and general radiologists without the use of AI-based CADe/x are different. An arbitration reading will be interpreted by these other radiologists specializing in breast imaging without the use of AI-based CADe/x. The details of the study interpretation design are shown in **Figure 1**.

The cutoff abnormality score of recall for test-positive reading with and without the use of AI-CADe/x will be defined using a 7-point malignant scale: 1 (definite normal), 2 (benign), 3 (probably benign), 4 (low suspicion of malignancy), 5 (moderate suspicion of malignancy), 6 (high suspicion of malignancy), and 7 (highly suggestive of malignancy). The cutoff abnormality score of stand-alone AI will be defined as > 10% of the scores (**Supplementary Data 1**).

Primary endpoint and secondary endpoint

The primary endpoint is the diagnostic accuracy of radiologists specializing in breast imaging with and without the use of AI-based CADe/x; this will be measured using the CDR, recall rate, sensitivity, positive predictive value (PPV), specificity, interval cancer rate, and AUROC considering the diagnostic data obtained within 12 months of interim analysis and within 24 months of final analysis from screening.

The secondary endpoints are diagnostic accuracy and differences in the following comparison groups with or without the use of AI-based CADe/x: between general radiologists with and without AI-based CADe/x; between radiologist arbitration reading for discordant cases (which involves discordant readings of breast and general radiologists without AI-based CADe/x) and breast radiologist with AI-based CADe/x; between radiologist arbitration reading for discordant cases (which involves discordant readings of breast and general radiologists without AI-based CADe/x) and general radiologists with AI-based CADe/x; between general radiologists with AI-based CADe/x and breast radiologists without AI-based CADe/x; between breast radiologists without AI-based CADe/x and stand-alone AI-based CADe/x; between general radiologists without AI-based CADe/x and stand-alone AI-based; between breast radiologists with AI-based CADe/x and general radiologists with AI-based CADe/x; between breast radiologists without AI-based CADe/x and general radiologists without AI-based CADe/x. The diagnostic accuracy will be assessed using cancer registry data as reference to calculate CDR, recall rate, sensitivity, PPV, specificity, interval cancer rate, and AUROC. The diagnostic accuracy of the comparison groups will be calculated using diagnostic data obtained within 12 months and 24 months of screening.

Sample size/power calculation

The sample size was estimated using McNemar's test to detect differences in CDR between groups of radiologists specializing in breast imaging with and without the use of AI-based CADe/x, with a 2-sided test at a significance level of 0.05% and 80% power. Sensitivity with and without AI-based CADe/x was 75.27% and 84.78%, respectively; the proportion of discordant pairs between with and without AI-based CADe/x for breast cancer screening was 12.1%, based on data from a previous retrospective study [12].

We assumed the cancer prevalence at 3.21 per 1,000 examinations [15]. With projection of the assumed cancer prevalence to the previous results, the target sample size was chosen as 32,714 participants, corresponding to approximately 16,000 participants per year.

Screening failure and drop-out criteria

Individuals who have provided informed consent but with breast parenchymal change due to surgery and mammoplasty or foreign substance as detected on mammography will be regarded as screening failure. Individuals can withdraw consent and discontinue participation at any time, even after enrollment, and the individuals who have withdrawn their informed consent will be regarded as drop-outs.

Data collection

The sources of information are as follows: 1) participant-reported demographics and medical history from participant information sheets and study questionnaires; 2) radiologist-collected mammographic findings from radiologist data collection forms; 3) standard AI findings from the image data file; 4) other mammographic and pathologic features from medical records and pathology reports, if available; and 5) breast cancer onset and survival status from the registry databases.

Mammograms from participants will be exported to the study platform (BEST image) for entry of reading results after digital imaging and communication in medicine standard de-identification according to the Health Insurance Portability and Accountability Act [9].

The Idea to Reality in Medicine (IRM)'s commercial BEST image is a certified local program included in the platform support for a multicenter study initiative by the Korean Society of Radiology.

Data management

IRM's commercial SNUPI program will be used for de-identification, including participant identification (ID), name, sex, birth date, and age; study instance unique identifiers; study date; series date; private tags; study description; and series description. A study participant ID will be assigned instead of the institution participant ID, and the relevant mapping information will be saved in an independent computer by the principal investigator at the site.

If the participant has mammography records at the relevant hospital from within the past 4 years, all such data will be exported to the study platform so that the reader can maintain the same environment as the actual reading at the hospital. Investigators or designated qualified staff in each institution will enter patient data obtained from mammography with or without the use of AI-based CADe/x. A valid username and password will be required to log into the eCRF system to secure patient data. Each patient will be automatically assigned a unique identification number at the time of enrollment according to the institution and registration order.

Data monitoring and cleaning

A data cleaning method will be employed to correct inconsistencies or errors not captured during data entry (e.g., outliers or conflicting data). Data queries will be identified on an ongoing basis during data collection. To enable evaluations and/or audits from regulatory authorities or the sponsor, the investigator agrees to retain all records, including the identity of all participants (sufficient information to link records, e.g., eCRFs and medical charts), source documents, detailed records of participant disposition, and adequate documentation of relevant correspondence (e.g., letters, meeting minutes, telephone call reports). A clinical study report will be submitted when the study is completed. All records and documents will be transferred to the document retention staff and retained for 5 years. If the investigator is unable, for any reason (e.g., retirement or relocation), to continue to retain study records

for the required period, the study records will be transferred to a designee, such as another investigator or another institution. After the retention period, records and documents will be destroyed to prevent disclosure of the contents of the documents and preserve confidentiality, and documents related to personal information will be destroyed according to the Personal Data Protection Act Enforcement Decree Article 16.

Statistical methods

Continuous data are described as standard deviations and quartiles. Categorical data are described as frequency and percentage (number, %). Missing data will be included as a separate category in some cases, depending on the nature of the variable. The statistical methods used in the descriptive analyses and statistical comparisons are presented in **Supplementary Data 2**. The intention-to-treat analysis will be performed primarily for all available endpoints; then, additional per-protocol analysis will be performed for the primary endpoint to verify that the results are consistent. An interim analysis is planned in 2026 and will include participants' demographics, clinical characteristics, and diagnostic accuracy in the primary and secondary study outcomes determined using cancer registry data as the gold standard reference data and obtained within 1 year of screening. Final analyses will be performed once all participant data have been collected and cleaned, and database lock has occurred. Certain subgroup analyses may be conducted depending on the sample size. Diagnostic accuracy endpoints (i.e., CDR, recall rate, sensitivity, PPV, specificity, and interval cancer rate) will be derived, and the corresponding 2-sided 95% confidence interval (CI) will be estimated using the Clopper-Pearson exact method. A logistic regression with the generalized estimating equation method and a multi-reader multicase ROC curve analysis will be performed to compare the diagnostic accuracy, considering the correlated nature of the design. For the reader-averaged AUROC, DeLong's method or bootstrap will be employed to estimate the 95% CI [16]. All statistical analyses will be performed using R version 4.0 (R Foundation, Vienna, Austria), and SAS version 9.4 or higher (SAS Institute, Cary, USA). Statistical significance will be set at $p < 0.05$. No correction will be performed for multiple comparisons.

DISCUSSION

We herein present the prospective multicenter study design of AI-STREAM using Lunit INSIGHT MMG AI-based CADe/x in a real-world population-based breast cancer screening program in Korea.

A few studies have shown that Lunit INSIGHT MMG AI-based CADe/x improves radiologists' performance and achieves an accuracy equivalent or superior to those obtained by radiologists alone [12,17].

After the retrospective cohort study's performance evaluation of the clinical value of a deep learning algorithm with small data, it is necessary to conduct a prospective study to confirm the reproducibility in the real clinical breast cancer screening environment. Since the main purpose of using AI in mammography is to increase the detection rate and reduce the recall rate of breast cancer in breast cancer screening, it should be performed in real breast cancer screening cohorts. AI-STREAM is designed to prospectively study a multicenter cohort and generate real-world evidence on the overall benefits and disadvantages of using AI-based CADe/x. In Korea, a single radiologist (either a radiologist specializing in breast imaging or a general radiologist) interpreted the mammogram. In this AI-STREAM study, the diagnostic

accuracy will be compared of radiologists specializing in breast imaging with and without the use of AI-based CADe/x for mammogram reading for breast cancer screening of women at average risk. The main purpose of this prospective study is to evaluate and confirm the changes in CDR and recall rate previously confirmed in a retrospective cohort study on a real breast cancer screening cohort. Thus, a multicenter prospective cohort study recruiting study subjects for 2 years is in progress, and a follow-up period of 2 years has been planned to evaluate interval cancers.

In many European studies, 2 readers interpreted mammograms. A recall occurs if either reader suggests it, through consensus, or after arbitration by a third readers. In the United States, mammograms are interpreted by a single reader, accompanied by CAD [18,19]. In this study, to evaluate various workflows using AI-based CADe/x for mammography interpretation, after the standard screening procedure in a single reading, double reading will be performed independently and retrospectively in other situations. The diagnostic accuracy in mammography readings between radiologists specializing in breast imaging and general radiologists with or without the use of AI-based CADe/x will be compared. The diagnostic accuracy of the arbitration reading of radiologists specializing in breast imaging will also be compared in various situations. Several studies have evaluated AI-CAD triaging mammography. Considering that the breast cancer rate is less than 1% in screening examinations, if AI-CADe/x can accurately identify cases in less time, with fewer resources, and without endangering the patient, we can reduce radiologists' workload and allocate more time on analyzing images with suspicious features and on any subsequent diagnostic workup [20,21]. AI-CADe/x used for triaging mammography examinations is advantageous because triaging negative examinations can spare radiologists' time and effort. Therefore, by reducing their overall workload and identifying cancers that have been missed by radiologists, AI-based CADe/x can act as a final consultant [13]. These studies may be of interest in European countries where there is a shortage of radiologists specializing in breast imaging, but with dual-reading systems for mammography screening examination [21,22].

If research results through prospective studies are secured, it will be possible to provide a solid basis for the positive results of actual deep learning algorithms for breast cancer screening.

In summary, AI-STREAM is a prospective multicenter cohort study that aims to generate real-world evidence on the overall benefits and disadvantages of using AI-based CADe/x for breast cancer detection. Patient enrollment is currently ongoing.

SUPPLEMENTARY MATERIALS

Supplementary Data 1
Collecting variables

Click here to view

Supplementary Data 2
Statistical data analysis

Click here to view

REFERENCES

1. Francies FZ, Hull R, Khanyile R, Dlamini Z. Breast cancer in low-middle income countries: abnormality in splicing and lack of targeted treatment options. Am J Cancer Res 2020;10:1568-91.
   PUBMED

2. National Cancer Center. Cancer Statistics. https://ncc.re.kr/main.ncc?uri=english/sub04_Statistics. Accessed September 22nd, 2021.

3. Kaplan HG, Malmgren JA, Atwood MK, Calip GS. Effect of treatment and mammography detection on breast cancer survival over time: 1990-2007. Cancer 2015;121:2553-61.
   PUBMED | CROSSREF

4. Choi KS, Yoon M, Song SH, Suh M, Park B, Jung KW, et al. Effect of mammography screening on stage at breast cancer diagnosis: results from the Korea National Cancer Screening Program. Sci Rep 2018;8:8882.
   PUBMED | CROSSREF

5. Lee K, Kim H, Lee JH, Jeong H, Shin SA, Han T, et al. Retrospective observation on contribution and limitations of screening for breast cancer with mammography in Korea: detection rate of breast cancer and incidence rate of interval cancer of the breast. BMC Womens Health 2016;16:72.
   PUBMED | CROSSREF

6. Majid AS, de Paredes ES, Doherty RD, Sharma NR, Salvador X. Missed breast carcinoma: pitfalls and pearls. Radiographics 2003;23:881-95.
   PUBMED | CROSSREF

7. Jo HM, Lee EH, Ko K, Kang BJ, Cha JH, Yi A, et al. Prevalence of women with dense breasts in Korea: results from a nationwide cross-sectional study. Cancer Res Treat 2019;51:1295-301.
   PUBMED | CROSSREF

8. Skaane P, Engedal K, Skjennald A. Interobserver variation in the interpretation of breast imaging. Comparison of mammography, ultrasonography, and both combined in the interpretation of palpable noncalcified breast masses. Acta Radiol 1997;38:497-502.
   PUBMED

9. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DS, Kerlikowske K, et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. Radiology 2017;283:49-58.
   PUBMED | CROSSREF

10. Giger ML, Chan HP, Boone J. Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. Med Phys 2008;35:5799-820.
    PUBMED | CROSSREF

11. Wu N, Phang J, Park J, Shen Y, Huang Z, Zorin M, et al. Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging 2020;39:1184-94.
    PUBMED | CROSSREF

12. Kim HE, Kim HH, Han BK, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. Lancet Digit Health 2020;2:e138-48.
    PUBMED | CROSSREF

13. Dembrower K, Wåhlin E, Liu Y, Salim M, Smith K, Lindholm P, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. Lancet Digit Health 2020;2:e468-74.
    PUBMED | CROSSREF

14. Sickles EA, D'Orsi CJ, Bassett LW. ACR BI-RADS® mammography. In: ACR BI-RADS® Atlas. Reston: American College of Radiology; 2013.

15. Hong S, Song SY, Park B, Suh M, Choi KS, Jung SE, et al. Effect of Digital mammography for breast cancer screening: a comparative study of more than 8 million Korean women. Radiology 2020;294:247-55.
    PUBMED | CROSSREF

16. Hillis SL. A marginal-mean ANOVA approach for analyzing multireader multicase radiological imaging data. Stat Med 2014;33:330-60.
    PUBMED | CROSSREF

17. Salim M, Wåhlin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. JAMA Oncol 2020;6:1581-8.
    PUBMED | CROSSREF

18. Taylor-Phillips S, Jenkinson D, Stinton C, Wallis MG, Dunn J, Clarke A. Double reading in breast cancer screening: cohort evaluation in the CO-OPS trial. Radiology 2018;287:749-57.
    PUBMED | CROSSREF

19. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 2015;175:1828-37.
PUBMED | CROSSREF

20. Lee EH, Kim KW, Kim YJ, Shin DR, Park YM, Lim HS, et al. Performance of screening mammography: a report of the alliance for breast cancer screening in Korea. Korean J Radiol 2016;17:489-96.
PUBMED | CROSSREF

21. Yoon JH, Kim EK. Deep learning-based artificial intelligence for mammography. Korean J Radiol 2021;22:1225-39.
PUBMED | CROSSREF

22. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Teuwen J, Broeders M, Gennaro G, et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. Eur Radiol 2019;29:4825-32.
PUBMED | CROSSREF