ISRCTN27361083 https://doi.org/10.1186/ISRCTN27361083

Evaluating artificial intelligence for detecting diseases on medical images and diagnostic tests

Submission date 16/12/2025	**Recruitment status** Recruiting	[_] Prospectively registered [_] Protocol [_] Statistical analysis plan [_] Results [_] Individual participant data [X] Record updated in last year
Registration date 17/04/2026	**Overall study status** Ongoing	
Last Edited 17/04/2026	**Condition category** Other	

Plain English summary of protocol

Background and study aims
Artificial intelligence (AI) is increasingly being used in hospitals to help identify diseases and abnormalities on medical scans such as X-rays and CT scans. Although many of these AI tools have been approved as safe to use in medical settings, there is very limited research on how well they actually work when used with real-world NHS patient data. This study aims to evaluate and compare the performance of AI algorithms designed to detect abnormalities on diagnostic images. By testing these AI tools on current patient data from NHS hospitals, we can determine whether they accurately identify or exclude abnormalities and whether they would be useful for doctors in their daily practice. This research is important because the NHS requires strong evidence before adopting new AI tools, and there is currently a gap in knowledge about how well these tools perform on UK patient data. This is a study platform that will serve as a framework to integrate studies whose study design is retrospective observational, data-only study, meaning no patients will be directly involved.

Who can participate?
The study will use anonymized medical images or investigations that were obtained as part of routine clinical care in NHS hospitals. Patients' data can be included unless they have opted out of NHS data sharing for research purposes through the National Data Opt-Out programme. No other inclusion or exclusion criteria apply, as this study does not involve direct patient participation.

What does the study involve?
The study will be conducted in NHS hospitals across the UK and will involve three main phases:
Phase 1 – Data Collection: Researchers will identify and collect anonymized clinical images or investigations from NHS hospitals. These images or tests will be from routine clinical care and will be carefully anonymized to remove all patient identifiable information.
Phase 2 – Establishing Ground Truth: Expert clinicians will independently review all the images or investigations to establish what the correct diagnosis should be for each test. If the experts disagree, a more experienced clinician will make a final decision. This creates the "ground truth" or reference standard against which the AI will be compared.
Phase 3 – AI Processing and Analysis: The AI software will analyze all the images or clinical investigation. The AI's diagnoses will then be compared against the expert clinicians decisions to

see how accurate the AI is at detecting normal and abnormal findings.
Phase 4 – Data Analysis: Researchers will calculate how well the AI performed, looking at measures such as sensitivity (how often it correctly identifies abnormalities) and specificity (how often it correctly identifies normal tests). The analysis will also examine whether the AI's performance differs across different patient groups, image qualities, and hospital settings.
No patient contact, additional imaging, or clinical procedures are involved, as this is entirely based on historical anonymized data.

What are the possible benefits and risks of participating?
Possible benefits: This research will help determine whether AI tools can safely and effectively support doctors in diagnosing medical conditions. If successful, this could lead to faster and more accurate diagnoses, reduced waiting times for patients, and less burden on NHS radiologists. The findings may also be used globally to help implement AI tools safely and effectively in other healthcare systems.
Possible risks: This is a retrospective study using anonymized historical data, so there are no direct risks to patients. Patients are not involved in any procedures or interventions. The only potential risk is if the AI tool performs poorly; However, this study is designed specifically to identify such issues before the tool would ever be used clinically. All data is strictly anonymized and protected under UK data protection laws (GDPR and Data Protection Act 2018).

Where is the study run from?
The study is coordinated by Oxford Clinical Artificial Intelligence Research (OxCAIR), Oxford University Hospitals NHS Foundation Trust. Data collection and analysis will take place at NHS hospitals across the UK, with participation from multiple hospital trusts in various regions.

When is the study starting and how long is it expected to run for?
January 2026 to January 2029

Who is funding the study?
Oxford Clinical Artificial Intelligence Research (OxCAIR) (UK)

Who is the main contact?
Dr Abdala Trinidad Espinosa Morgado, abdala.espinosa@ouh.nhs.uk

Contact information

Type(s)
Principal investigator

Contact name
Prof Alex Novak

ORCID ID
https://orcid.org/0000-0002-5880-8235

Contact details
OxCAIR, John Radcliffe Hospital, Headley Way
Oxford
United Kingdom

OX3 9DU
+44 (0)1865 221499
alex.novak@ouh.nhs.uk

Type(s)
Scientific, Public

Contact name
Dr Abdala T Espinosa Morgado

ORCID ID
https://orcid.org/0000-0003-0967-3554

Contact details
OxCAIR, John Radcliffe Hospital, Headley Way
Oxford
United Kingdom
OX3 9DU
+44 (0)1865 221499
abdala.espinosa@ouh.nhs.uk

Additional identifiers

Integrated Research Application System (IRAS)
363971

Study information

Scientific Title
Systematic assessment of the medical utility of radiology and diagnostic artificial intelligence - retrospective analysis

Acronym
SAMURAI-Retro

Study objectives
Primary objective:
To assess the performance/accuracy of AI algorithms in detecting key pathologies on anonymised medical diagnostic tests against a reference standard (ground truth).

Secondary objectives:
1. To evaluate factors which may affect the accuracy/performance of AI algorithms, including patient factors (age, sex, ethnicity), diagnostic test factors (image quality, acquisition parameters), and pathology factors (size, severity, subtype)
2. If multiple AI algorithms are evaluated, to compare the diagnostic accuracy of algorithms to assess if there is a statistically significant difference in performance between AI tools

Exploratory objectives:
1. To systematically explore diagnostic accuracy across different imaging modalities, clinical contexts, and patient/pathology subgroups

2. To evaluate factors affecting AI performance including image quality and technical characteristics
3. To investigate AI algorithm performance across different patient demographics and assess for algorithmic bias or equity issues
4. To analyse cost-effectiveness and resource implications
5. To evaluate reliability and consistency of ground truth determination
6. To generate high-quality anonymised datasets for continuous AI quality assurance and regulatory benchmarking

Ethics approval required
Ethics approval required

Ethics approval(s)
Submitted 16/01/2026, NHS Research Ethics Committee (REC), Oxford University Hospitals NHS Foundation Trust (Headley Way, Headington, Oxford, OX3 9DU, United Kingdom; +44 (0)1865 572240; debbie.franklin@ouh.nhs.uk), ref: 363971

Primary study design
Observational

Secondary study design
Retrospective observational study

Study type(s)

Health condition(s) or problem(s) studied
Evaluation of artificial intelligence algorithms for detecting various pathologies across multiple diagnostic modalities including medical imaging (X-ray, CT, MRI, ultrasound) and other diagnostic tools (e.g., electrocardiograms). Specific conditions will be defined at the sub-study level.

Interventions
Data Collection: Anonymised medical diagnostic test datasets (imaging and other diagnostic investigations) collected from routine clinical care via Electronic Patient Records (EPR) and clinical IT systems

Ground Truth Establishment: Reference standard determined within each sub-study (e.g., through sub-specialist consultant reports or expert arbitration methodology using two independent experts with arbitrator)

AI Algorithm Application: CE-approved or late-stage development AI algorithms applied to anonymised datasets, either locally or via secure data transfer to vendors

Performance Analysis: AI outputs compared against ground truth reference standard to calculate diagnostic accuracy metrics (sensitivity, specificity, accuracy, area under the curve, positive predictive value, negative predictive value)

Statistical Analysis: Statistical tests applied to identify differences between AI algorithms and across subgroups (using methods such as McNemar's test, Cochran's Q test, one-way ANOVA)

Data Handling:
Full compliance with GDPR and Data Protection Act 2018

Advanced pseudonymisation/anonymisation techniques applied
Secure, password-protected, encrypted databases
National Data Opt-Out respected

Intervention Type
Other

Primary outcome(s)
1. The accuracy/performance of AI algorithms in detecting key pathology on anonymised medical diagnostic tests measured using sensitivity, specificity, overall accuracy, area under the curve (AUC), positive predictive value (PPV), negative predictive value (NPV) at not applicable (retrospective data analysis)

Key secondary outcome(s)

Completion date
01/01/2029

Eligibility

Key inclusion criteria
Since this is a data-only retrospective study, there are no direct participant inclusion criteria. Rather, the inclusion criteria apply to the imaging/diagnostic data:
1. Anonymised medical diagnostic investigations (imaging or diagnostic tests) obtained as part of routine clinical care
2. Diagnostic data that can be anonymised without compromising data integrity
3. Images/tests meeting quality standards for AI algorithm analysis

Healthy volunteers allowed
Yes

Age group
Mixed

Lower age limit
2 Years

Upper age limit
120 Years

Sex
All

Total final enrolment
0

Key exclusion criteria
1. Imaging or diagnostic investigations where data cannot be anonymised or where anonymisation compromises data integrity
2. National Data Opt-Out: Data from patients listed in the National Data Opt-Out database who

have formally opted out of having their data shared for research purposes
3. Substudy-specific exclusions: Additional exclusion criteria (e.g., age restrictions, specific imaging modality requirements, or pathology-specific criteria) that are dependent on individual substudies

Date of first enrolment
01/03/2026

Date of final enrolment
31/12/2028

Locations

Countries of recruitment
United Kingdom

England

Study participating centre
Oxford University Hospitals
John Radcliffe Hospital
Headley Way
Headington
Oxford
England
OX3 9DU

Sponsor information

Organisation
Oxford University Hospitals NHS Trust

ROR
https://ror.org/03h2bh287

Funder(s)

Funder type

Funder Name
Oxford Clinical Artificial Intelligence Research (OxCAIR)

Results and Publications

Individual participant data (IPD) sharing plan

IPD sharing plan summary
Not expected to be made available