

## A Research Study to Investigate the Use of Artificial Intelligence to Improve Breast Cancer Screening

### **What is artificial intelligence?**

“Artificial Intelligence”, commonly known as “AI”, is a broad term for computer software programmes that aim to imitate human intelligence by learning and making decisions.

Generally, these software programmes or “algorithms” consist of a set of instructions to perform a specific task. However, some tasks, which may be easy for humans to do, are very difficult to write as instructions for computers to complete. A classic example is deciding if an image shows a cat or a dog.

This is where AI algorithms can be useful as, rather than writing step-by-step instructions for the task, these algorithms are “trained” by being given an input (such as a picture of a cat) and being told the correct output (“this is a picture of a cat”). For each training image the algorithms’ own output is compared with the correct output and small adjustments are made to its internal calculations to minimise the difference between the two.

With a large amount of good quality training data, these algorithms learn to complete a task well without ever being explicitly told how to do so.

### **Artificial intelligence and cancer**

In radiology, AI algorithms are trained in this way using thousands of medical images which a radiologist has marked as containing cancer or not. A new set of images, which were not used for training, can then be presented to test how well the algorithm detects cancer.

### **What is this study specifically trying to do?**

Our research will test how well different AI algorithms detect breast cancer in mammograms, with a view to improving breast cancer screening.

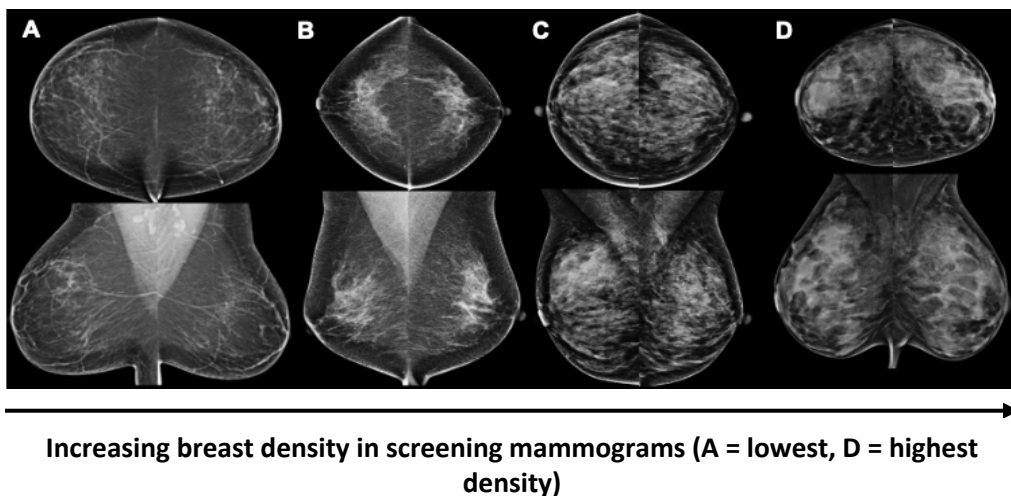
This work is part of a Cancer Research UK programme grant for ‘Risk adaptive breast cancer screening: a tailored imaging approach’.

### **How might artificial intelligence improve breast cancer screening?**

Breast screening (using mammogram x-rays) in the UK is conducted as part of a National Health Service (NHS) Breast Screening Programme for women aged between 50 and 70 years. Women over the age of 70 are able to self-refer. At certain centers women age 47 to 73 are also invited as part of a trial. Screening and early detection of breast cancer is known to improve outcomes for the women affected, however it is a labour-intensive process. At present, screening mammograms are read by two expert readers, and around 70-80% of the cancers are detected.

It is believed that AI algorithms have the potential to work alongside expert readers to ease the workload in the breast screening programme and make it more effective by:

- Improving the accuracy of reading mammograms and therefore cancer detection
- Acting as one of the two readers to ease the workload of readers
- Making reading more efficient by marking and prioritising mammograms which the algorithm believes are more likely to contain cancer
- Identifying women with dense breast tissue, for whom mammography is less sensitive, and who would benefit from additional imaging methods such as ultrasound or MRI.



### What does the study involve?

As part of this research, we plan to build a database that will eventually contain mammograms from approximately 150,000 women from the Cambridge and Norwich Breast Screening Programmes. The database will include different types of cases to be representative of the case proportions (normal cases and cancer cases) in the screening programme.

Initially we are collecting the existing mammograms of women who attended screening between 2011 and 2020. During the course of our research, we will continue to collect new screening mammograms until the end of 2022.

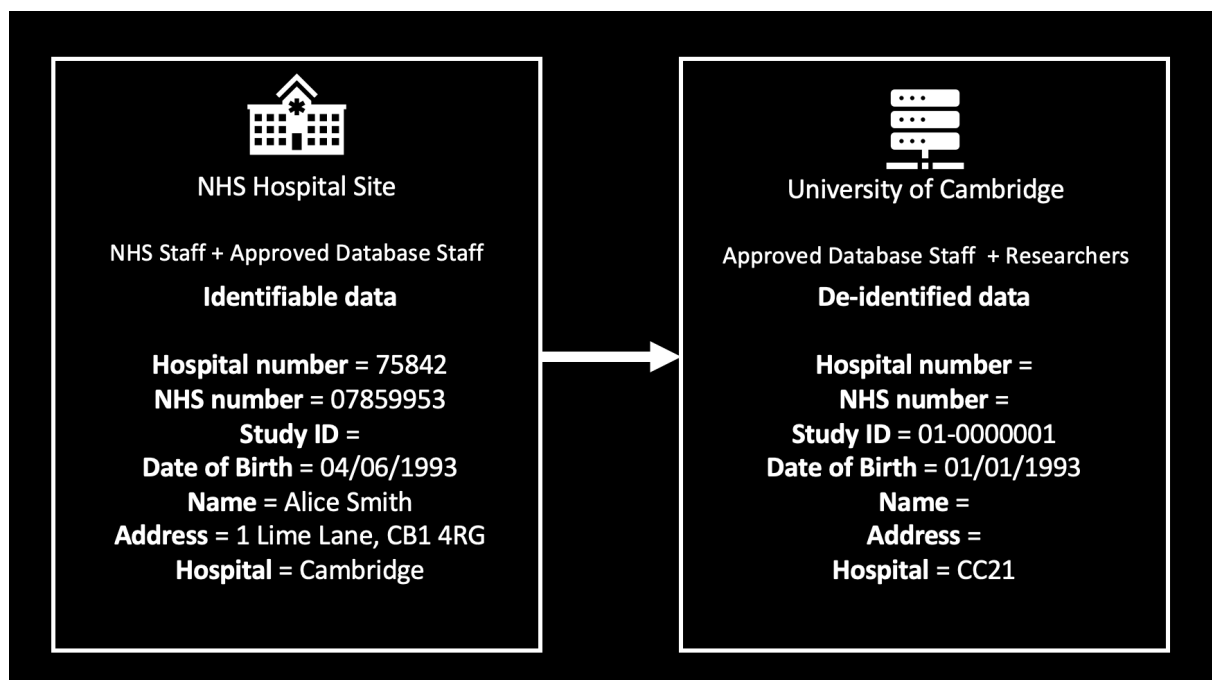
### The key database with identifiable data

First, a key database, of **identifiable (non-anonymised)** data, will be created and stored at the hospital (Cambridge University Hospitals NHS Foundation Trust / Norfolk and Norwich University Hospital NHS Foundation Trust) securely within the NHS IT system. This key database will:

- contain information that is identifiable, which is normally collected as part of routine care (date of birth, NHS number, hospital number, screening date, screening site, image accession number, episode record ID and screening number). These identifiers allow us to find the mammogram images and clinical information relating to the image, for each woman, such as; details of previous breast biopsies, treatment (e.g. surgery or radiotherapy), and genetic risk factors
- contain a new unique study ID, to identify each case
- be managed according to the security and data management rules for confidential information within the health sector
- not be used for testing and will remain within the hospital site
- be accessed by specific members of the research team who have been granted permission for access.

### The final testing database with de-identified data

Second, a final testing database containing only **de-identified** information will be created and stored at the University of Cambridge. This database will be the same as the key database except all the information that could be used to identify someone (name, address, date of birth, NHS number, etc) will be removed. The testing database will contain the de-identified mammogram images, clinical information relating to the image and unique study ID for each woman. It is this de-identified database that will then be used for testing the AI algorithms.



De-identification process from the key to testing database

How will the different AI algorithms be tested and assessed?

21.10.21

V1.5

Many AI algorithms have been developed by universities and computer software experts following recent technological advances. A number of the most advanced of these AI algorithms will be tested on the de-identified final testing database.

Our work will also aim to highlight the performance and limitations of each AI algorithm, which is vital for clinicians to understand how to use them effectively in future real-world practice.

### **Who will monitor the access and use of the database?**

The key database and the final testing database will be governed by a Database Access Committee who will oversee the use of data from both databases.

The de-identified final testing database for research will be stored at University of Cambridge. All of our studies testing the AI algorithms and the analysis of the results will take place at the University of Cambridge.

In certain circumstances, limited amounts of de-identified data may be shared outside the UK with academics and commercial health companies under strict restrictions on how that data can be accessed and used. For example, one possible reason for data sharing would be to enable companies to adapt an algorithm to our mammography machine images, as they may not have had access to this type of machine before.

The testing database will also be a valuable resource of a large number of high-quality images for creating, developing and testing new AI algorithms.

To be able to use AI algorithms within the NHS, regular monitoring and data security infrastructure is required. This requires a partnership between the NHS, academics and commercial health companies to develop, test, regularly monitor and update systems, to ensure both patient safety and patient data security.

The national ethics boards, and authorities which approved this study, have concluded that consent from individual women is not required for their mammograms to be de-identified and included in the creation of an image database governed by strict information governance policies.

### **Opting-out of your data being used**

If you do not want your mammograms and clinical information to be used for this research, or you would like more information about this study, please contact our Breast Research Nurses.

#### **Breast Research Nurses contact details:**

Phone: 01223 348937

Email: [braid.trial@nhs.net](mailto:braid.trial@nhs.net)

Address:

IRAS 258761

21.10.21

V1.5

Department of Radiology  
University of Cambridge School of Clinical Medicine  
Box 218 Cambridge Biomedical Campus  
Cambridge  
CB2 0QQ

Website: <https://radiology.medschl.cam.ac.uk/cc-media/>