# OPPORTUNITY OVERVIEW

- OPTIMAM is a large, curated and centralized database of mammogram images that is collected from several NHS breast screening programme sites across the UK

- The data currently spans 10+ years and is continually updated with new images and associated clinical data from existing screening sites as well as newly onboarded sites

- Unprocessed and processed 2D & 3D mammography images are collected of normal, benign and malignant cases with rich associated clinical data relating to screen detected and interval cancers

- 2D images are marked with clinical region of interest and expert-determined ground truth annotations

- The resource has been designed to be shared for research and commercial purposes, including in Virtual Clinical Trials, Computer Aided Detection, Artificial Intelligence, Image Perception Studies and Training and Quality Assurance

- The ongoing collection of images and is funded by CRUK and the database is hosted and managed by staff at the Royal Surrey NHS Foundation Trust

- We are accepting applications from commercial or non-profit organisations, healthcare institutions and academic centres for access to the database

- Please see the OMI-DB (OPTIMAM Mammography Image Database) website for further information

USE CASES
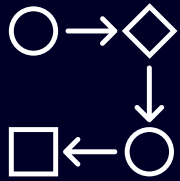
Training and Testing of Deep Learning Computer Aided Detection Algorithms

Considerably aided by the presence of a large number of well-labelled malignant cases as well as details of interval cancers and large numbers of normal cases
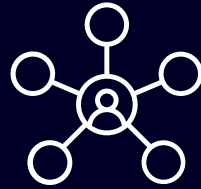
Training of Models Designed to Predict Risk and Stratify Patient Populations

Supported by the existence of sequential screening episodes and known outcomes over a long time period (9+ years)
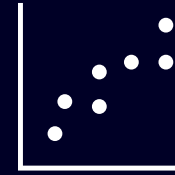
Density Investigation & Fundamental Physics of Image formation

Technical repeats and calculated breast density make investigations into density and image quality possible, alongside unprocessed data to explore fundamental physics of imaging

Pre-Market Validation Exercises

Companies benefit from using a UK dataset to understand how well their algorithm performs prior to UK launch and inform potential further development

Example Use Case: Lancet Digital Health publication *(2020) in which OMI-DB was used in the development of an AI algorithm for diagnosis of breast cancer in mammography*

VALUABLE DATA POINTS

ASSOCIATED CLINICAL DATA

- Invasive status

- Disease grade

- Mutation status

- Screen Reader opinion

- Lesion description

- Size of tumour

- Hormone replacement status

- *For full details please see description of contents*

ANNOTATED IMAGES

- The majority of images within the database have been annotated by expert radiologists

- Expert radiologists from the collection sites retrospectively use information from the Electronic Patient Record (EPR) and screening system to identify the tumour and draw a rectangular ROI (region of interest) around its boundaries representing ground truth

- The coordinates and various properties of the tumour are recorded and described in the associated image data

VALUABLE DATA POINTS

LONGITUDINAL DATA

- Longitudinal data points create a wealth of research applications

- Information on previous screening events and interval cancers, as well as the associated data from assessment, biopsy and tomosynthesis imaging are available within the database and can be longitudinally linked

- New mammography images and data are continually added to OMI-DB adding to the richness of longitudinal data available

UNPROCESSED IMAGES

- Considerable amounts of unprocessed images enable research into the fundamental physics of medical imaging

- Special arrangements have been made at image collection sites so that unprocessed images are not discarded, whereas most PACS systems used in the NHS only store processed images

VALUABLE DATA POINTS

3D TOMOSYNTHESIS IMAGES

- >2,000 tomosynthesis cases are available for sharing, of which 845 are malignant cases

- Images are derived through follow-up tomosynthesis conducted when women are recalled for further assessment and are therefore typically of a single breast or one view of the breasts

- The accompanying 2D screening images for these 3D tomosynthesis images are annotated

BREAST DENSITY

- In addition to the clinical information, information regarding breast density can also be included upon request

- Breast density is calculated by Volpara Density (a fully automated breast density assessment software)

DATABASE SIZE – 2D IMAGES

The database currently contains mammography images from 173,319 women collected since 2011

A breakdown of the 2D dataset available for sharing is listed below:

Type	# Cases
Normal	10,000
Malignant (marked)*	5,500
Malignant (unmarked)	800
Benign (marked)*	600
Benign (unmarked)	1,000
Interval	1,000

Bespoke requests can be accommodated e.g. 50% of the cases or an increased number of normal cases

Malignant and Benign cases are sub-classified into marked and unmarked, depending on whether they have been annotated by an expert mammography reader

Interval Cases are those that were detected in the 3-year interval between screening events

DATABASE SIZE – 3D IMAGES

A breakdown of the 3D tomosynthesis dataset available for sharing is listed below:

Type	# Cases
Normal*	1000
Malignant (Marked)	500
Malignant (Unmarked)	345
Benign** (Marked)	65
Benign (Unmarked)	250
Interval	70

Images are derived through follow-up tomosynthesis conducted when women are recalled for further assessment and are therefore typically of a single breast or one view of the breasts

The accompanying 2D screening images for the 3D tomosynthesis cases are annotated

Normal cases include those recalled for assessment and recalled for assessment plus proven normal with biopsy

** Benign cases are proven with biopsy*

EQUIPMENT MANUFACTURER

The current equipment manufacturer distribution within OMI-DB is as follows:

Manufacturer	Proportion of Total Images within Database
Hologic	91.59%
GE Healthcare	4.94%
Siemens	0.72%
Philips	0.19%
Others	2.56%

Expansion of data collection and further onboarding of new screening centres will achieve a wider diversity of X-ray equipment manufacturers that are used for breast screening across the UK

SUMMARY & NEXT STEPS

- Please find further details about the database at the Royal Surrey NHS Foundation Trust [OMI-DB website](#)

- OPTIMAM Mammography Image Database (OMI-DB) is available for licencing, please submit applications to the [online form](#)

- The web form will alert the team of your interest, and they will send out a more detailed application form to be completed in full which will be reviewed by the [OPTIMAM Steering Committee](#)

- Access is subject to a sharing agreement with Cancer Research Horizons which have the right to commercialise the intellectual property of OMI-DB

CANCER
RESEARCH
HORIZONS

FURTHER FASTER TOGETHER

THANK YOU

For further information, please contact:
Eleanor Stead

eleanor.stead@cancer.org.uk