OPTIMAM Database

Data Licensing Opportunity

Clinically linked mammography dataset

FURTHER FASTER TOGETHER
We are beating cancer

Overview

Indication: Breast
Outcome: Normal to Malignant
Geography: UK

~470,000
Clients

~7 million
Images

~1.3 million
Studies

Data Types

- 2D images
- 3D images
- Clinical
- Pathological
- Ethnicity
- Deprivation

See slide 10 for further information on how to access data

Background

The OPTIMAM Database (OMI-DB) is a large, centralised & de-identified database of **mammography images** collected from **multiple NHS breast screening sites** across the **UK**

- OMI-DB has been designed to help overcome **the lack of large, curated & representative medical image databases**

- The ongoing collection of images & management of the database is conducted by the **Royal Surrey NHS Foundation Trust** & is funded with support from **Cancer Research UK (CRUK)**

- OMI-DB collects both **processed & unprocessed mammogram images**, associated **clinical, surgical** & **pathological** data for **all cases,** as well as expert-determined region of interest **(ROI) annotations** for a subset of cases

- Where available, follow-up assessment via **3D screening (digital breast tomosynthesis)** assessment & magnetic resonance imaging (MRI) of very high-risk cases is also collected

Case Description

The database includes both **2D** (FFDM) and **3D** (digital breast tomosynthesis) images
- Majority of cases are 2D: as per UK breast cancer screening guidelines, 3D images are only collected when a case is recalled for further assessment due to a suspicious finding

2D

- **4 images** per screening event (2 views of each breast - medio-lateral oblique & cranio-caudal)

- Each case may include images from **multiple screening events**

- **3-year intervals** between screening events

3D

- DBT images are typically taken **2-4 weeks** after the initial screening event

- Number of DBT images **varies per case**: some may have 2 views of each breast, the majority are **single breast views**

- Each DBT case will include the **2D images** from **prior screening events** for that client

- DBT images **do not** have expert annotations

Note: One Case refers to all images & data associated with one client

Definitions

Normal	Cases not classified as malignant or benign, or interval cancer
Benign	Cases that are classified as **benign** by either surgery, or biopsy, or previously classified as benign
Malignant	Cases that are classified as **malignant** by either surgery, or biopsy, or previously classified as malignant
Interval	Symptomatic cases which were **detected & classified as malignant** after a previous screening appointment at which a client received a normal result & **before their next scheduled screening** appointment, or previously classified as interval cancer
Marked vs. Unmarked	Cases are sub-classified into marked & unmarked, depending on whether **region of interest** (ROIs) have been annotated by an **expert mammography reader**

Key differentiators

Large volume

- Imaging & clinical data from **~470,00 clients** across the UK

- **~7 million images** collected via variety of mammography systems

Longitudinal data

- Data collection spans **10+ years** & new mammography images & data are **continually added**

- Information on **previous screening episodes** & **interval cases** are available & can be longitudinally linked

In-depth clinical data & annotation

- Each case is associated with **detailed clinical information**, including biopsy, clinical, surgery & lesion details

- **Region of interest (ROI) annotations** by expert radiologists are available for a number of cases

Standardised & curated

- Data has been **anonymised & standardised** by an expert database management team

- **Bespoke data requests** can be accommodated

Available data: Summary

CANCER RESEARCH HORIZONS

2D images from **~470,000** cases

Outcome	Number of Cases Available
Normal	424,339
Benign	11,029
Malignant	15,404
Interval Cancers	2,860
Unknown	15,727

3D images from **~10,200** cases

Outcome	Number of Cases Available
Normal	4,796
Benign	1,457
Malignant	3,912
Interval Cancers	49

Note: One Case refers to all images and data associated with one client. For more detailed statistics on available data, please visit the OPTIMAM website

Available data: Summary

8,675 marked cases available

Outcome	Number of Cases Available
Benign Marked	1,432
Malignant Marked	7,160
Prior to Interval Cancer Marked	83

*Note: Marked = ROIs have been annotated by an expert mammography reader; annotations are available on **2D images only***

Data Description

Annotated Images
- **Expert radiologists** at the screening sites retrospectively **identify lesions** on the mammogram images & draw a rectangular ROI around its boundaries representing **ground truth** (bounding box)
- Coordinates & various properties of the tumour/lesions are recorded & described in the **associated image data**

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

Associated Clinical Data
- Includes detailed **biopsy, clinical, surgery & lesion** information e.g. invasive status, disease grade, mutation status, screen reader opinion, tumour size, HER2 receptor score etc.
- Please reach out to CRH to request **full details**

Longitudinal Data
- New mammography images & data are **continually added** to OMI-DB from existing sites & **additional NHS sites** which are being onboarded
- Information on **previous screening episodes & interval cancers** are available within the database & can be longitudinally linked

Technical & Biological Features
- Cases are recorded & classified with any technical or biological features such as **microcalcifications** or image blurring that can enable further investigation and research
- **Breast density data** (calculated by Volpara Density) may be requested (sharing is subject to further approval)

Applying for access

Data access requests

- To access OMI-DB, companies are required to **submit a data access request** to the database team via the online form
  - The web form will alert the team of your interest & they will send out a more detailed application form to be completed, which will be reviewed by the OPTIMAM Steering Committee

Key information required:

- Description of the proposed study & intended use of the data (in both technical & lay terms)

- The specific data types that you would require access to

- Description of the potential patient benefit generated from the use of the dataset

Licence agreements

Access is subject to a licence agreement with Cancer Research Horizons & Royal Surrey NHS Foundation Trust, which includes terms that align with CRH's Guiding Principles to Commercial Data Partnerships

Key terms:

- Data access is offered on a **non-exclusive** basis & solely for the **approved purpose**

- Any **IP** generated from use of the data will be **owned by the company***

- Licences are **time-limited** (typical term is 3-years)

Data access will be facilitated through a Google Cloud Storage Bucket

For licencing inquiries please reach out the CRH team

Thank you

For further information, please contact:

Jakob Kridl
jakob.kridl@cancer.org.uk

CANCER RESEARCH HORIZONS

S:CORT
Stratification in COloRecTal cancer

FURTHER FASTER TOGETHER
We are beating cancer