Clinical Imaging 118 (2025) 110369

Contents lists available at ScienceDirect

Clinical Imaging

journal homepage: www.elsevier.com/locate/clinimag

Breast Imaging

Diversity, inclusivity and traceability of mammography datasets used in development of Artificial Intelligence technologies: a systematic review

Elinor Laws [a,b,c], Joanne Palmer [a,b,c], Joseph Alderman [a,b,c], Ojasvi Sharma [a,b], Victoria Ngai [d], Thomas Salisbury [e], Gulmeena Hussain [a], Sumiya Ahmed [f], Gagandeep Sachdeva [g], Sonam Vadera [a,h], Bilal Mateen [i,j,k], Rubeta Matin [l], Stephanie Kuku [m], Melanie Calvert [c,n,o,p,q], Jacqui Gath [r], Darren Treanor [s,t,u,ac], Melissa McCradden [v,w], Maxine Mackintosh [x,y], Judy Gichoya [z], Hari Trivedi [z], Alastair K. Denniston [a,b,c,aa,ab], Xiaoxuan Liu [a,b,c,ab,*]

[a] *University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK*
[b] *Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK*
[c] *National Institute for Health and Care Research Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK*
[d] *University College London Medical School, London, UK*
[e] *South Tyneside and Sunderland NHS Foundation Trust, Sunderland, UK*
[f] *University of Leicester Medical School, Leicester, UK*
[g] *The Royal Wolverhampton NHS Trust, UK*
[h] *University Hospitals of Leicester NHS Trust, Leicester, UK*
[i] *University College London, Institute of Health Informatics, London, UK*
[j] *PATH, Seattle, Washington, United States*
[k] *Wellcome Trust, London, UK*
[l] *Oxford University Hospitals NHS Foundation Trust, Oxford, UK*
[m] *University College London, London, UK*
[n] *Birmingham Health Partners Centre for Regulatory Science and Innovation, University of Birmingham, Birmingham, UK*
[o] *Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK.*
[p] *National Institute for Health and Care Research Applied Research Collaboration West Midlands, University of Birmingham, Birmingham, UK*
[q] *National Institute for Health and Care Research Blood and Transplant Research Unit in Precision Transplant and Cellular Therapeutics, University of Birmingham, Birmingham, UK*
[r] *Independent Cancer Patients' Voice, London, UK*
[s] *Leeds Teaching Hospitals NHS Trust, Leeds, UK*
[t] *University of Leeds, Leeds, UK*
[u] *Department of Clinical Pathology, Linköping University, Linköping, Sweden*
[v] *Department of Bioethics, The Hospital for Sick Children, Toronto, Canada*
[w] *Genetics & Genome Biology, SickKids Research Institute, Toronto, Canada*
[x] *Genomics England Limited, London, UK*
[y] *The Alan Turing Institute, London, UK*
[z] *Department of Radiology, Emory University, USA*
[aa] *National Institute for Health and Care Research Biomedical Research Centre, Moorfields Eye Hospital/University College London, London, UK*
[ab] *Birmingham Health Partners Centre for Regulatory Science and Innovation, University of Birmingham, Birmingham, UK*
[ac] *Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden*

A R T I C L E I N F O

Keywords
Artificial intelligence
Mammography

A B S T R A C T

Purpose: There are many radiological datasets for breast cancer, some which have supported the development of AI medical devices for breast cancer screening and image classification. This review aims to identify mammography datasets (including digitised screen film mammography, 2D digital mammography and digital breast tomosynthesis) used in the development of AI technologies and present their characteristics, including their transparency of documentation, content, populations included and accessibility.

* Corresponding author at: University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK.
E-mail address: x.liu.8@bham.ac.uk (X. Liu).

https://doi.org/10.1016/j.clinimag.2024.110369
Received 16 September 2024; Received in revised form 4 November 2024; Accepted 22 November 2024
Available online 26 November 2024

Materials and methods: MEDLINE and Google Dataset searches identified studies describing AI technology development and referencing breast imaging datasets up to June 2024. The characteristics of each dataset are summarised. In particular, the accompanying documentation was reviewed with a focus on diversity and inclusion of populations represented within each dataset.

Results: 254 datasets were referenced in the literature search, 190 were privately held, 36 had barriers which prevented access, and 28 were accessible. Most datasets originated from Europe, East Asia and North America. There was poor reporting of individuals' attributes: 32 (12 %) datasets reported race or ethnicity; 76 (30 %) reported female/male categories with only one dataset explicitly defining whether these categories represented sex or gender attributes.

Conclusion: Through this review, we demonstrate gaps in the data landscape for mammography, highlighting poor representation globally. To ensure datasets in breast imaging have maximum utility for researchers, their characteristics should be documented and limitations of datasets, such as their representativeness of populations and settings, should inform scientific efforts to translate data-driven insights into technologies and discoveries.

1. Introduction

Breast cancer is the most common neoplastic disease worldwide with around 2.3 million people diagnosed globally each year.[1] Many settings have implemented mammography-based breast screening programs with the aim of detecting asymptomatic cases allowing earlier treatment.[2] Breast screening programs reduce mortality caused by breast cancer but they also present significant resource implications. For example, in the UK 2.2 million mammograms are performed annually: the majority of which are normal but all require two expert mammography reader opinions and occasionally, an additional arbitration process to confirm the report.[3] There is an incentive to develop strategies to reduce resource implication of breast cancer detection and treatment.[4]

Artificial Intelligence (AI) has shown promise in breast cancer detection, in some cases matching human mammography readers in experimental settings and in other cases, enhancing accuracy when used in conjunction with readers.[5,6] It is hoped that if applied to a screening setting, AI may relieve some of the workload burden, saving radiologists' time and facilitating provision of services in less resourced settings.[7] As of June 2024, the United States Food and Drug Administration has approved 39 mammography Software as Medical Device (SaMD) products.[8]

Developing safe and equitable AI technologies requires training and testing on large amounts of data. A recent review into mammography datasets demonstrated that there are relatively few that are open access.[9] When choosing an appropriate dataset, there are many factors to consider including the completeness, traceability and representativeness of the data.[10,11] Recent systematic reviews of open access dataset documentation in dermatology, ophthalmology and Covid-19 have shown a lack of representativeness across geographical areas and demographic groups even where disease burden is high, as well as lack of clinically relevant population group labels.[12–14] The concern is that where datasets used for developing AI technologies are not representative of the target population, they may underperform for underrepresented groups and often, it is groups already experiencing health inequity that are underrepresented.[12,15] There is value in providing detailed dataset description within accompanying publications regardless of whether the data are open access. Detailed data description can allow readers to make sound judgements on the potential impacts on associated AI technologies, to understand the extent to which ethical considerations may have been made, and whether relevant population groups are represented.

This study aims to review the landscape of datasets for mammography used in training and testing of AI technologies. In particular, the study aims to assess the level of inclusion and documentation of demographic groups known to have worse outcomes from breast cancer or participation in breast screening programs (including minoritized ethnic groups, transgender/gender non-conforming groups).[4,16,17] This review is part of the STANDING Together initiative - STANdards for data Diversity, INclusivity and Generalisability (www.datadiversity.org).[18,19]

2. Materials and methods

2.1. Search strategy and selection criteria

This systematic review aimed to identify datasets used in the development of AI technologies through a search of bibliographic databases and dataset search engines (Google Datasets Search). Searches were conducted on 04/06/2024 using MEDLINE via Ovid to identify articles that referred to mammography datasets in AI research. The full search strategy is provided in Appendix 1. Only English Language texts were included. A PRISMA flowchart is provided in Fig. 1. Given this systematic review is not directly relating to health outcomes, it did not qualify for PROSPERO registration. The search terms used are as follows:

((Dataset) OR (Database) OR (Collection) OR (Repository)) AND ((Artificial Intelligence) OR (AI) OR (Natural Language Processing) OR (NLP) OR (Machine Learning) OR (Support Vector Machine) OR (Neural Network) OR (Deep Learning)) AND ((Adenocarcinoma) OR (Ductal Carcinoma) OR (Lobular Carcinoma) OR (Triple Negative) OR (Angiosarcoma) OR (Phyllodes Tumour)) AND ((Breast) OR (Breast Cancer)) AND ((Diagnosis) OR (Prediction) OR (Screening)) AND ((Mammogram) OR (Imaging) OR (Image Interpretation, Computer Assisted) OR (Imaging Processing, Computer Assisted) OR (Magnetic Resonance Imaging) OR (MRI) OR (Ultrasound)).

2.2. Identification of datasets

Abstracts were independently screened by JP, VN, EL and OS using Rayyan.[20] Full texts were screened by EL, JP, OS, SA, TS or GH to identify the name and access requirements for referenced mammography datasets. All datasets with mammography images were included, even if multi-modal. A Google Dataset search for 'breast cancer dataset' was also conducted by GS and EL and reviewed by JA. Access requirements were classified into: Group 1 - open access; Group 2 - managed access (defined as datasets that required formal application, approvals or payment for access); Group 3 - managed access with barriers (defined as datasets that were reported as managed access but had practical barriers to access like out-of-date contact details); and Group 4 - private (not reported as accessible). Where access required communication with curators (e.g. email correspondence or submission of an application form), we defined an acceptable response period as four weeks; if there was no response within that time, the dataset was classed as Group 3 - managed access with barriers. Datasets with digitised screen film mammography, 2D digital mammography and digital breast tomosynthesis were included. Whilst the terms 'magnetic resonance imaging' (MRI) and 'ultrasound' were within the search, associated datasets were only included if they also involved mammography images. Datasets composed entirely of another included dataset were excluded as duplicates, as were datasets containing only MRI, ultrasound imaging, histopathology, text/numerical or non-human data and no mammography data. Reviews were also excluded as the purpose of the study was

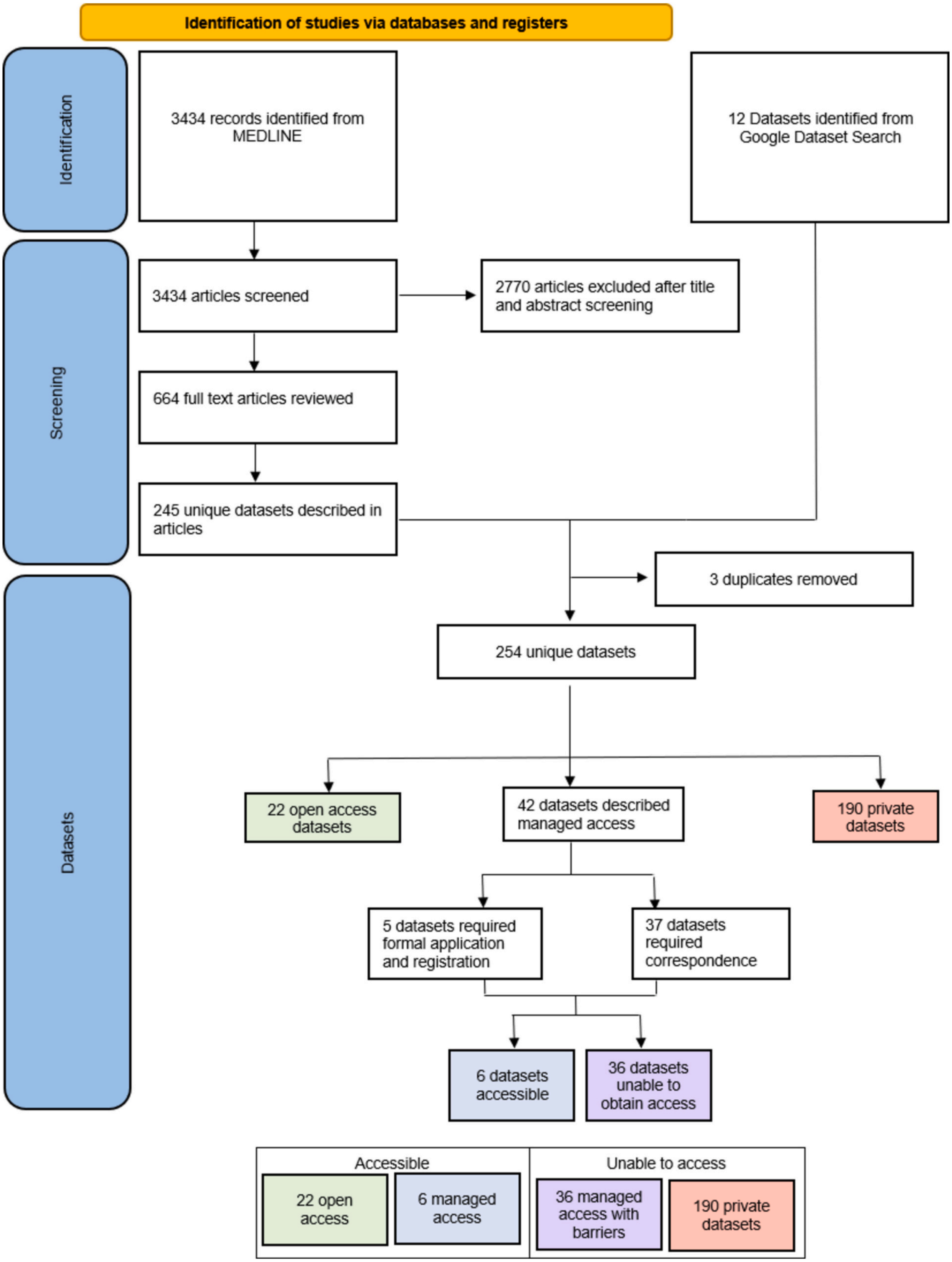

Fig. 1. PRISMA Flowchart showing the process of dataset identification
Articles identified through MEDLINE search were reviewed. Referenced datasets were assessed for accessibility and categorised into: Group 1 - open access; Group 2 - managed access (defined as datasets that required formal application, approvals or payment for access); Group 3 - managed access with barriers (defined as datasets that were reported as managed access but had practical barriers to access like out-of-date contact details); and Group 4 - private (not reported as accessible). 254 datasets were identified with 22 open access and six managed access datasets being accessible. The additional 226 datasets were not accessible and only corresponding information written in the research articles could be reviewed.

to identify mammography datasets that had been used in the development of AI models.

2.3. Extraction of dataset characteristics

Information regarding format of dataset documentation, accessibility, version history, content (population, geographical location, data collection setting, modifications to the data), and metadata (labels, clinical data, file format) were extracted, (Appendix 2) in line with the STANDING Together recommendations.[21] Where dataset information was hosted in multiple internet locations, for feasibility, we extended our search of formation to no more than three successive web pages. All data extraction was completed by EL, OS and XL. Information was considered 'reported' if it was available in the dataset documentation, dataset description, host site or publication. Information was considered reported if it was present at either the individual participant or aggregate level. If documentation stated a valid reason for not collecting a particular parameter (for example, where a demographic attribute could not be collected for governance reasons), this was also classed as reported. For all open access datasets, a sample of mammography images were inspected for embedded metadata using Horos online DICOM viewer (Version 3.3.6 for Mac, Horos Project). For the managed access datasets, clarification around data availability was sought in direct correspondence with dataset owners/custodians through online meetings. For private datasets, data could not be accessed and therefore our review was limited to the documentation within corresponding papers identified during the search.

2.4. Patient and public involvement and engagement (PPIE)

A PPIE committee were consulted during each phase of the development of this manuscript from funding application to problem definition and manuscript review. During quarterly meetings, the research team presented project updates and PPIE members contributed their perspectives on how the project should be conducted. For example, the list of demographic attributes to include within the dataset extraction phase was decided by the PPIE committee. A PPIE partner also formed part of the project working group who reviewed the manuscript.

3. Results

3.1. Datasets identified from the literature search

The MEDLINE search identified 3434 articles. After abstract screening, 2770 were excluded as they did not reference AI, machine learning, or mammography. In total, 664 full-text papers were reviewed which identified 245 unique datasets. A further nine unique datasets were identified as part of the Google dataset review. Two additional datasets that were commonly cited (Mini-MIAS and CBIS-DDSM) were excluded from full review as they were derived from datasets that were already included (MIAS and DDSM). No datasets identified, to our knowledge, included synthetic data. We include all citations for papers referencing mammography datasets that we were unable to access in Appendix 1.3. AI technologies described in the papers identified included image segmentation models, breast mass detection and classification, calcification pattern recognition and breast cancer prediction.

3.2. Dataset accessibility

Of the 254 included datasets, 22 were open access where data could be freely downloaded and reviewed. Six managed access datasets required a data access request to the data. These 28 accessible datasets are listed in Table 1. Two additional datasets (Clínica Chavarría's 2020 mammogram target dataset and Breast Screen Victoria dataset) requiring correspondence with curators replied to our requests for access but were unable to share data due to local governance restrictions. An additional two managed access datasets did not reply to our requests for access. Of the 226 referenced datasets which could not be accessed, we classified 36 as 'managed access with barriers' and 190 as 'private' - we refer to these datasets collectively as 'unable to access'.

3.3. Dataset characteristics

In the 254 identified datasets, there are over 12 million mammograms from over three million individuals from 34 countries. These figures are likely an underestimation as 134 datasets (including DDSM, MIAS, SNUBH-MDB and Mammograms Breast-Cancer Images) described only the number of individuals or number of images, and 18 reported neither. In total, 186 (74 %) datasets reported geographical origin of the data (Fig. 2). The UK's OMI-DB dataset represents over two million of those images and >170,000 individuals. The number of images included in the datasets ranged from 10 images to two million.

A total of 33 datasets (13 %) included images originally acquired on film and converted to digital format. A total of 103 datasets (41 %) were reported to be derived from a screening programme, 10 (4 %) were derived from 'diagnostic settings' where symptomatic individuals undergo mammography as part of the diagnostic process, 19 (7 %) were derived from both screening and diagnostic settings and 120 (47 %) did not report where the data were derived from. One dataset was derived from a clinical trial and another two datasets contained 'phantom'/'augmented' images based on diagnostic mammograms. In total, 106 datasets (42 %) described whether consent was obtained from individuals or not.

3.4. Reporting of individuals' attributes

Overall, 113 (45 %) datasets reported age, 32 (13 %) reported race or ethnicity and 76 (30 %) reported sex or gender. Of the 28 accessible datasets, only four reported race or ethnicity. The number of datasets reporting individuals' attributes and additional metadata reporting categories are presented in Fig. 3.

Definitions for race and ethnicity were not provided within any dataset documentation, although clarification was sought during the interaction with the data curation team of the OMI-DB dataset, where groups were mapped to the United Kingdom National Health Service Data Model and Dictionary Ethnicity Categories.[22] Where race/ethnicity were reported, there was significant variation in categories used (Appendix 1.4). The numbers of individuals represented within each race or ethnicity category was variable between datasets. 12 of 32 datasets that included race or ethnicity reported that over 70 % of individuals were within the 'White' or 'Caucasian' category. However, 7 of 32 datasets that included race or ethnicity, and were derived from the United States of America, reported similar numbers of individuals within different racial or ethnic categories such as 'White' and 'African American'.

Of the 254 identified datasets, 76 included information on 'female'/'male' categorisation. For the majority of datasets, it was unclear whether 'male' and 'female' categories were referring to sex assigned at birth, or gender. Through direct communication with the dataset curation teams of OMI-DB we were able to clarify that the categories within their datasets referred to sex assigned at birth, and that separate data on gender identity was not available - this clarification was not possible for the other datasets. One dataset included information about individuals who identify as a gender that is different to their sex assigned at birth. Two accessible datasets reported 'male' individuals: BCDR included one 'male' individual (and 5059 'female'); BCS-DBT included 13 'male' individuals (and 1721 'female'). Only five datasets (2 %) reported socioeconomic status of individuals. Appendix 1.5 displays the full list of metadata categories reported for each accessible dataset.

Table 1
Characteristics of accessible datasets.
OMI-DB: OPTIMAM Mammography Imaging Database, DDSM: Digital Database for Screening Mammography, MIAS: Mammographic Image Analysis Society Digital Mammogram Database, BCS-DBT: Breast cancer screening—digital breast tomosynthesis, CDD-CESM: Categorised Digital Database for Low energy and Subtracted Contrast Enhanced Spectral Mammography, SNUBH-MDB: The Seoul National University Bundang Hospital Digital Mammographic database, CBIS-DDSM: Curated Breast Imaging Subset of Digital Database for Screening Mammography, DMID: Digital mammography Dataset for Breast Cancer Diagnosis Research, SSBS: Scottish Breast Screening Service.
Country of origin: country from which the data originated.
Source of diagnostic label: how were the diagnoses (benign, malignant) determined? i.e. through pathology report (biopsy), from a mammography report written by a radiology expert, from an AI model classification.
*Imaging data were not directly transferred but correspondence with dataset curators regarding data availability and metadata categories was had.

Search	Dataset	Number of citations	Accessibility	Year of dataset completion	Image type	Number of patients	Number of images	Country of origin	Source	Source of diagnostic label
MEDLINE	DDSM[47] (Associated dataset: CBIS-DDSM[48])	192	Open access	1999	Thumbnails within website	2620	Not reported	USA	Screening	Biopsy, report by radiology expert
	INbreast[49]	75	Open access	2010	.tgz file	115	410	Portugal	Both screening and diagnostic	Biopsy, report by radiology expert
	MIAS[50] (Associated dataset: Mini-MIAS[51])	102	Open access	1994	.pgm	Not reported	322	UK	Not reported	Not stated
	Breast Cancer Digital Repository[52]	16	Open access	2012	Images and labels visible on website	1010	3703	Portugal	Not reported	Report by radiology expert
	OMI-DB[53]	11	Regulated access	2016–present	.dcm	Over 170,000 (collection ongoing)	Over 2,000,000 (collection ongoing)	UK	Both screening and diagnostic	Report by radiology expert, Label generated by AI model, Associated clinical data e.g. biopsy, derived data and annotated ROI
	BCS-DBT[54]	3	Open access	2018	.dcm	5060	22,032	USA	Screening	Report by radiology expert
	EMBED_Open_Data[55]	3	Regulated access	2021	.png	22,382	480,323	USA	Both screening and diagnostic	Biopsy, report by radiology expert
	The Chinese Mammography Database (CMMD)[56]	3	Open access	2023	.dcm	1775	3728	China	Not Reported	Biopsy
	Breast Microcalcifications dataset[57]	1	Open access	2020	.jpg	100	200	Not stated	Screening	Biopsy, report by radiology expert
	CDD-CESM[58]	1	Open access	2021	.jpg	326	2006	Egypt	Not reported	Report by radiology expert
	SNUBH-MDB[59]	1	Open access	2015	.dcm	Not reported	49	Not stated	Not reported	Not stated
	DMID[60]	1	Open access	2023	.dcm	Not reported	510	India	Screening	Report by radiology expert
	VinDr[61]	1	Open access	2022	.dcm	Not reported	5000	Vietnam	Both	Report by radiology expert, biopsy
	Mammographic Mass[62]	1	Open access	2022	.data	Not reported	benign: 516; malignant: 445	Germany	Screening	Not reported
	VICTRE[63]	1	Open access	2022	.dcm	2986 ‘subjects’	217,913	Not stated	Representative of a screening population’	Label generated by AI model

(continued on next page)

Table 1 (*continued*)

Search	Dataset	Number of citations	Accessibility	Year of dataset completion	Image type	Number of patients	Number of images	Country of origin	Source	Source of diagnostic label
	Transformers Improve Breast Cancer Diagnosis from Unregistered Multi-View Mammograms[64]	1	Regulated access	2022	.dcm	1775	3728	Not stated	Not reported	Not reported
	Ambra UNIFESP Mammography dataset[65]	1	Regulated access	2024	.dcm	100	941	Brazil	Screening	Reported by radiology expert
	CSAW-S[66]	1	Regulated access	2024	.png	172	Not reported	Not stated	Not Reported	Reported by radiology expert
	*SBSS[67]	1	Regulated access	2023	.dcm	55,916	Not reported	UK	Screening	Reported by radiology expert, biopsy
Google dataset search	KAU-BCMD[68]		Open access	2021	.dcm	1416	5662	Saudi Arabia	Not reported	Report by radiology expert
	Mammograms-Breast Cancer Images[69]		Open access	2019	.jpg	Not reported	10	Not stated	Not reported	Not stated
	RSNA Mammography Breast Cancer Detection PNG[70]		Open access	Not reported	Not reported	Not reported	Not reported	Not stated	Not reported	Not reported
	Full mammogram[71]		Open access	Not reported	Not reported	Not reported	Not reported	Not stated	Not reported	Not reported
	Mammogram Mastery: A Robust Dataset for Breast Cancer Detection and Medical Education[72]		Open access	April 2024	JPEG (.jpg)	Not reported	745 original images, 9685 augmented images	Iraq	Not reported	Report by radiology expert
	BIRAD Mammo[73]		Open access	Not reported	JPEG (.jpg)	Not reported	Not reported	Not stated	Not reported	Not reported
	Breast Micro-Calcifications Dataset with Precisely Annotated Sequential Mammograms[57]		Open access	2021	.jpg and .dcm	100	200	Cyprus	Not reported	Report by radiology expert
	Breast Mammography Image Dataset with Masses[74]		Open access	2023	.png	Not reported	2347 original images from DDSM, MIAS and Inbreast used to create 24,576 synthetic images	Not stated	Not reported	Not reported
	Mammogram Density Assessment Dataset[75]		Open access	2024	JPEG (.jpg)	Not reported	Not reported	Not stated	Not reported	Report by radiology expert

3.5. Traceability

Within the 664 full text articles reviewed, the datasets that were most accessible were also the datasets that were most cited. The DDSM dataset published in 1999 (and its associated dataset, CBIS-DDSM) was most frequently cited ($n = 210$). Number of citations increased with the age of the datasets and older datasets were consistently cited up until the year of literature search (2024), even where new and updated versions of the dataset were available. Older accessible datasets (DDSM and MIAS) more commonly had versions with multiple host sites, some of which were created by users rather than the original curators of the datasets and in many cases, dataset documentation did not accompany data on alternative host sites. Some datasets with different names consisted of related data: for example, the DDSM dataset led to the creation of CBIS-DDSM which was a subset of images modified into a condensed format. Names and access links of dataset versions were referenced interchangeably between research papers.[23–25] Authors of papers using the accessible datasets often referenced ‘open access’ links that were invalid; overall, there were 419 citations of the accessible mammography datasets identified by MEDLINE search, but only 147 of these provided valid access links. Of the 226 datasets that were not accessible as part of our review, only 67 reported a name or any identifier for the datasets despite many referencing previous use of the data in different research projects.

4. Discussion

This review provides an overview of mammography datasets, and summarises their accessibility, content, and documentation. We particularly focused on dataset composition and documentation. Effective dataset documentation can give insight into diversity and inclusion and data shifts, which can all have implications on model generalizability, fairness and post-market surveillance considerations.

Of the 254 unique datasets identified, we found relatively few that

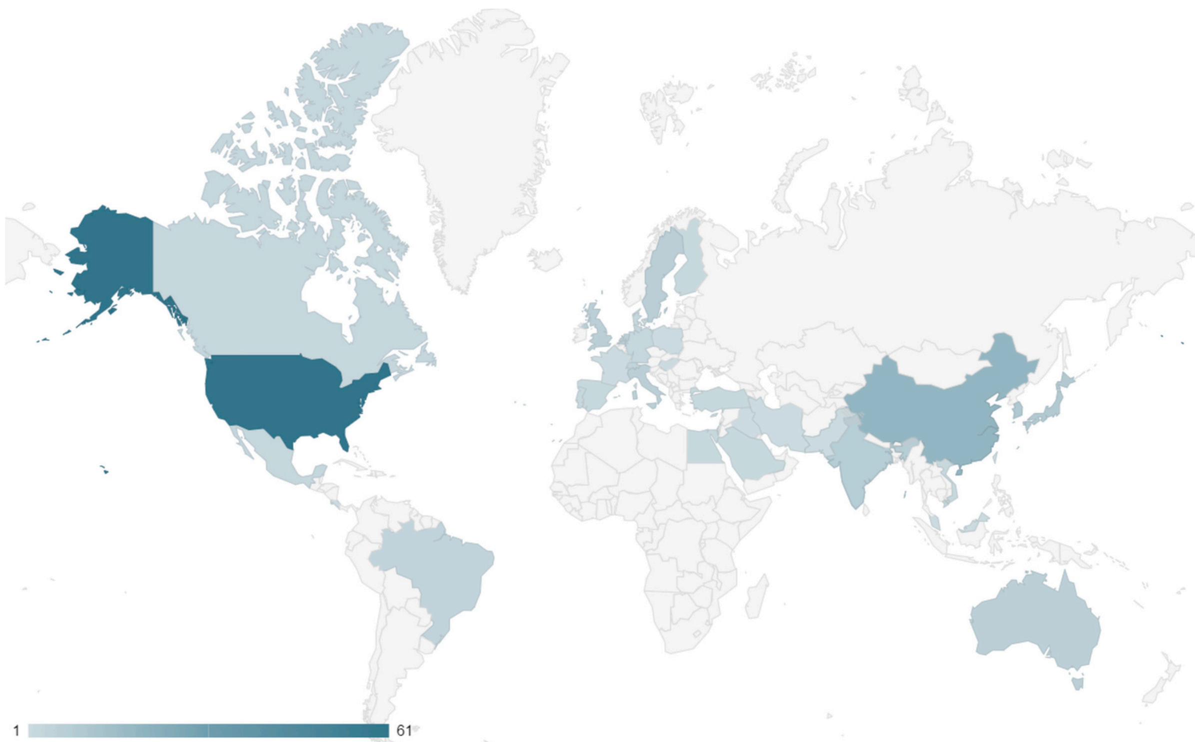

Fig. 2. A figure to show the geographical origin of the datasets identified as part of this review.
Number of datasets per country: Not reported (n = 190), USA (n = 61), China (n = 19), South Korea (n = 10), Japan (n = 9), India (n = 8), Sweden (n = 7), Australia (n = 6), Netherlands (n = 6), UK (n = 6), Italy (n = 6), Cyprus (n = 2), Portugal (n = 2), Saudi Arabia (n = 2), Spain (n = 2), Taiwan (n = 5), Brazil (n = 4), Switzerland (n = 4), Canada (n = 3), Germany (n = 3), Israel (n = 3), Mexico (n = 3), Denmark (n = 2), Egypt (n = 2), Finland (n = 2), Turkey (n = 2), Vietnam (n = 2), Costa Rica, France, Hungary, Iran, Iraq, Malaysia, Pakistan, Poland. Some papers referenced several datasets from different countries, and5some datasets were made up of a collection of data from several different countries.

were open access: around two thirds had no public access at all (neither open, nor managed access); and out of the 63 datasets that claimed to provide access (open or managed), only 28 were fully and practically accessible within the timelines described in this manuscript. This has important implications for the medical data science community as we found these 28 accessible datasets to be most commonly cited and therefore likely to disproportionately over-represent and over-influence data-driven innovations in breast cancer. For example, only four of the 28 accessible datasets reported race or ethnicity categories which limits the ability to incorporate these parameters within the development of any derived AI technology. The limitations of these open access and highly cited datasets are likely to be repeated and potentially amplified at scale through their use. We note four key areas of concern within datasets used for the development of AI in mammography: (a) the paucity of reporting of individuals' demographic attributes; (b) diversity and inclusivity of datasets in use; (c) barriers to dataset access; and (d) traceability of the datasets used to develop AI technologies.

4.1. Demographic reporting

Individuals' personal attributes are recognised to be predictors of disease outcome and are therefore particularly important to consider in the context of AI in healthcare. In our review, 45 % of datasets reported age, 30 % of datasets reported 'female'/'male' categories, only 13 % datasets reported race or ethnicity and 2 % datasets reported socioeconomic status, even though studies have shown that these demographic attributes are associated with disparate breast cancer outcomes.[17,26] Additionally, rarely were demographic labels defined within the dataset documentation even though definitions and categories of demographic attributes can vary between countries.[27] Within our review, race and ethnicity categories varied greatly between datasets. The variability in racial and ethnic categories reported is not surprising given that they are social constructs that are dependent on local socio-political structures which are themselves highly variable.[28]

Only one dataset specifically referred to those who identify as a gender different to their sex assigned at birth.[31] The lack of reporting gender identity within the context of mammography is particularly notable as people who identify as a gender different to the sex they were assigned at birth are at risk of breast cancer: for example, trans-women and non-binary people whose sex assigned at birth was male, and who are taking long term oestrogen therapy, are at increased risk of breast cancer, and trans-men and non-binary people assigned female at birth are at risk of breast cancer, even where breast reduction surgery has been performed.[16] Lack of representation within datasets risks groups, such as the transgender community, being systematically excluded from benefits of data-driven insights and innovations.[16]

Demographic data such as race, ethnicity, sex, gender and socioeconomic status, are key to understanding where health disparities exist, allowing data-driven technologies to be built with inclusivity and equity in mind.[32,33] However, an important caveat is that collecting demographic data can itself be a cause of stigmatisation leading to worse healthcare and inequitable outcomes, and can present risk of re-identification.[34–36] There is work to be done in earning trust from communities about the value of collecting demographic data, whilst also recognising the right for privacy, and it is essential to hear the perspectives of the communities themselves.[37] There is also a need to educate the medical and health research workforce to improve the capacity for recording complex demographic data responsibly.[32]

4.2. Diversity and inclusion

Another consideration for data used to develop AI technologies is

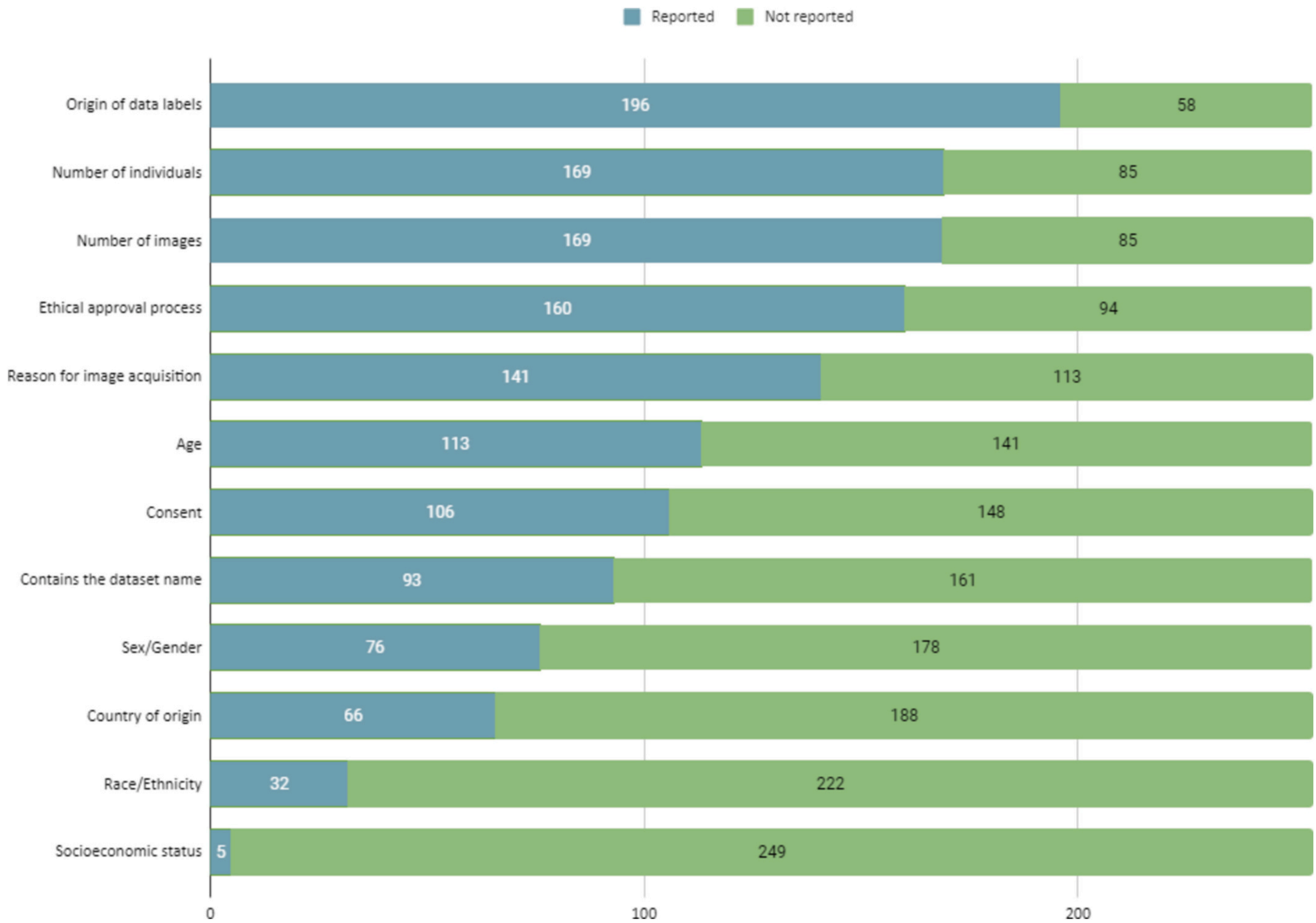

Fig. 3. The number of datasets reporting metadata categories (n = 254).

> Sex is defined as a label assigned at birth (either intersex, female or male) based on biological characteristics (e.g. reproductive anatomy, genetics, physiology), whereas gender refers to a person's psychological, emotional, social, and overall sense of identity which classifies them as woman, man or neither.[29] Race is a social construct used as a proxy for human variation, often based on skin tone and other physical characteristics.[28] Ethnicity is complex to define but broadly refers to a person's identity relating to their origin and background.[30]

dataset diversity. Breast cancer is a disease that affects people globally yet the majority of the global south were not represented within our review and there were no datasets representing any African country, except for Egypt. The Human Development Index (HDI) – a statistical index with values from 0 to 1 used to classify countries according to life expectancy, income and education within populations – is associated with survival from breast cancer and many African countries are within the lowest HDI categories.[38] There is a risk that countries already experiencing inequity in breast cancer outcomes may also be excluded from digital health innovation as a result of a lack of data that represents them.[39] Disparate breast cancer outcomes are also seen within higher HDI settings: in the US, minoritized racial/ethnic groups and those with lower socioeconomic status are known to have worse breast cancer outcomes.[17,26] Whilst some datasets from the USA identified within our review had similar representation between African American and White populations, many of the mammography datasets show under-representation of different racial or ethnic groups.[17] If mammography data are under-representative of diverse population groups, the resulting AI technologies are less likely to work well for these communities which could contribute to an exacerbation of health inequity.[40]

4.3. Barriers to access

Despite 254 datasets being identified within this review, only 28 were practically accessible. Many datasets were reported as accessible from the corresponding author of the associated publication but rarely were replies received, and other dataset host sites were outdated which meant access and support was limited. Where contact was obtained, we could seek clarification via video-call and email and this was valuable in contributing to our understanding of the data. For example, we could ask for the curators' definitions of demographic labels such as 'sex' and 'ethnicity.' There is a global initiative that aims to make data findable, accessible, interoperable, and reusable (FAIR) but many datasets identified did not meet these criteria.[41]

The advantages of open access data are improved availability and accessibility, serving as a resource for the scientific and innovation community. However, there are legitimate trade-offs compared to a managed data access model, including increased data privacy risk, reduced data governance and oversight, and increased cost burden with no direct value-return for the data curator. There is a risk that countries with fewer resources to invest in curation of large managed access datasets may be left behind. Additionally, the managed access process is time consuming and resource intensive for data users, which may act as a disincentive where open access alternatives are available – in our review the open access datasets had significantly higher citation counts than the managed access datasets, even when accounting for dataset age.[42] Our experience interacting with a few managed access dataset curation teams were highly valuable but if this model is to be successful, there needs to be appropriate investment and resources to provide this level of support.[43,44] It is also important to acknowledge the sensitive nature of health data, and careful decision making should be undertaken to justify whether making data accessible is ethical and necessary.

4.4. Dataset traceability

Within our review, we noted that authors referenced multiple dataset host sites and used different dataset names (intended to refer to different versions of the dataset) interchangeably. Inconsistent dataset referencing makes it difficult to trace back the data underpinning a particular AI technology in turn limiting ability to assess its appropriateness for different population groups. It is also important to consider the context of the dataset from a technical point of view: 33 datasets included within our review were derived from film screening images that had been digitised. A large proportion of screening worldwide is now conducted with fully digitised technology, putting the applicability of these datasets into question.

4.5. Limitations

Whilst this systematic review used a robust search strategy to identify literature referencing mammography datasets, the extracted list is unlikely to be comprehensive of all available mammography datasets. Firstly, there is a delay between dataset publishing and reporting within research articles, meaning mammography datasets published after June 2024 are not included.[45,46] Secondly, our search strategy included a Google dataset search but did not include direct searches of alternative dataset repositories such as Kaggle, Mendeley and Github.[12,13]

With regards to the datasets identified as part of our review, we made reasonable attempts to gain access wherever possible by contacting corresponding authors of papers and applying through official dataset access processes. However, a significant proportion of identified datasets were not accessible as part of this review and therefore only limited dataset review could be conducted and additional metadata may be available such as demographic data like race, ethnicity, sex and gender. We were unable to review the completeness of any dataset, in part due to time restraints but also due to lack of reporting. We also chose to limit our search to English language texts which will have excluded a number of articles written in additional languages. Finally, we were not able to review all available images due to time constraints and there may be some embedded dataset documentation that we have not accessed.

4.6. Conclusion

A key step to creating AI health technologies that are inclusive and equitable is to aim for representation of the full diversity of the population who could benefit from such technologies. Noting that some demographic information can be highly sensitive and, in some contexts, may put individuals at risk, it is essential that the collection and use of this data is undertaken with the support of the individuals and communities represented within those datasets. The setup of managed access, well maintained mammography datasets should be considered with appropriate investment and focus on lower income settings. Ultimately, availability of high quality and inclusive health data is on the critical path to creating more effective, safe and trustworthy AI technologies for all.

CRediT authorship contribution statement

Elinor Laws: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Joanne Palmer:** Writing – review & editing, Data curation. **Joseph Alderman:** Writing – review & editing, Visualization, Supervision, Methodology. **Ojasvi Sharma:** Project administration, Investigation, Formal analysis, Data curation. **Victoria Ngai:** Data curation. **Thomas Salisbury:** Writing – review & editing, Data curation. **Gulmeena Hussain:** Data curation. **Sumiya Ahmed:** Data curation. **Gagandeep Sachdeva:** Data curation. **Sonam Vadera:** Writing – review & editing. **Bilal Mateen:** Writing – review & editing. **Rubeta Matin:** Writing – review & editing. **Stephanie Kuku:** Writing – review & editing. **Melanie Calvert:** Writing – review & editing. **Jacqui Gath:** Writing – review & editing. **Darren Treanor:** Writing – review & editing. **Melissa McCradden:** Writing – review & editing. **Maxine Mackintosh:** Writing – review & editing. **Judy Gichoya:** Writing – review & editing. **Hari Trivedi:** Writing – review & editing. **Alastair K. Denniston:** Writing – review & editing, Supervision, Methodology. **Xiaoxuan Liu:** Writing – review & editing, Supervision, Resources, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

MJC is Director of the Birmingham Health Partners Centre for Regulatory Science and Innovation, Director of the Centre for Patient Reported Outcomes Research, and is a National Institute for Health and Care Research (NIHR) Senior Investigator. MJC receives funding from the NIHR, UK Research and Innovation (UKRI), NIHR Birmingham Biomedical Research Centre, NIHR, Applied Research Collaboration (ARC) West Midlands, Research England, European Regional Development Fund, and the NIHR Blood and Transplant Research Unit in Precision Transplant and Cellular Therapeutics. MJC also receives funding from Innovate UK (part of UKRI), Macmillan Cancer Support, UCB Pharma, GSK, Anthony Nolan, Gilead Sciences, European Commission, European Federation of Pharmaceutical Industries and Associations, and The Brain Tumor Charity. MJC has received personal fees from Aparito, CIS Oncology, Gilead, Halfloop, Takeda Pharmaceuticals, Merck, Daiichi Sankyo, Glaukos, GSK, the Patient-Centered Outcomes Research Institute, Pfizer, Genentech, and Vertex Pharmaceuticals, outside of the submitted work.

MMackintosh is an employee of Genomics England and Founder and Director of One HealthTech and Data Science for Health Equity (within One HealthTech Ltd).

JG is a grantee of the 2022 Robert Wood Johnson Foundation Harold Amos Medical Faculty Development Program and declares support from RSNA Health Disparities Grant (EIHD2204), Lacuna Fund (67), Gordon and Betty Moore Foundation, and NIH (NIBIB) MIDRC grant under contracts 75N92020C00008 and 75N92020C00021.

XL has received consulting fees from Hardian Health and was previously a Health Studies Scientist at Apple.

Appendix 1. MEDLINE search strategy

#	Search Terms for Pubmed/Medline Version 2	Results
1	Dataset (“dataset”[Publication Type] OR “datasets as topic”[MeSH Terms] OR “dataset”[All Fields])	150,421
2	Database (“database”[All Fields] OR “database s”[All Fields] OR “databased”[All Fields] OR “databases”[All Fields] OR “databasing”[All Fields])	846,919
3	Collection (“collect”[All Fields] OR “collectable”[All Fields] OR “collected”[All Fields] OR “collecting”[All Fields] OR “collection”[All Fields] OR “collections”[All Fields] OR “collects”[All Fields])	1,686,349
4	Repository (“repositories”[All Fields] OR “repository”[All Fields])	23,455
5	1 OR 2 OR 3 OR 4	2,542,668
6	Artificial Intelligence OR AI (“artificial intelligence”[MeSH Terms] OR (“artificial”[All Fields] AND “intelligence”[All Fields]) OR “artificial intelligence”[All Fields] OR (“antagonists and inhibitors”[MeSH Subheading] OR (“antagonists”[All Fields] AND “inhibitors”[All Fields]) OR “antagonists and inhibitors”[All Fields] OR “ai”[All Fields])	1,330,544
7	Machine Learning (“machine learning”[MeSH Terms] OR (“machine”[All Fields] AND “learning”[All Fields]) OR “machine learning”[All Fields])	152,539
8	Support Vector Machine (“support vector machine”[MeSH Terms] OR (“support”[All Fields] AND “vector”[All Fields] AND “machine”[All Fields]) OR “support vector machine”[All Fields])	30,698
9	Neural network (“neural networks, computer”[MeSH Terms] OR (“neural”[All Fields] AND “networks”[All Fields] AND “computer”[All Fields]) OR “computer neural networks”[All Fields] OR (“neural”[All Fields] AND “network”[All Fields]) OR “neural network”[All Fields])	157,140
10	Deep learning (“deep learning”[MeSH Terms] OR (“deep”[All Fields] AND “learning”[All Fields]) OR “deep learning”[All Fields])	79,403
11	6 OR 7 OR 8 OR 9 OR 10	1,487,169
12	adenocarcinoma (“adenocarcinoma”[MeSH Terms] OR “adenocarcinoma”[All Fields] OR “adenocarcinomas”[All Fields] OR “adenocarcinoma s”[All Fields])	534,082
13	Ductal carcinoma (“carcinoma, ductal”[MeSH Terms] OR (“carcinoma”[All Fields] AND “ductal”[All Fields]) OR “ductal carcinoma”[All Fields] OR (“ductal”[All Fields] AND “carcinoma”[All Fields])	47,244
14	Lobular carcinoma (“carcinoma, lobular”[MeSH Terms] OR (“carcinoma”[All Fields] AND “lobular”[All Fields]) OR “lobular carcinoma”[All Fields] OR (“lobular”[All Fields] AND “carcinoma”[All Fields])	10,293
15	Triple negative ((“triple”[All Fields] OR “triples”[All Fields]) AND (“negative”[All Fields] OR “negatively”[All Fields] OR “negatives”[All Fields] OR “negativities”[All Fields] OR “negativity”[All Fields])	31,450
16	Angiosarcoma (“haemangiosarcoma”[All Fields] OR “hemangiosarcoma”[MeSH Terms] OR “hemangiosarcoma”[All Fields] OR “haemangiosarcomas”[All Fields] OR “angiosarcoma”[All Fields] OR “angiosarcomas”[All Fields] OR “hemangiosarcomas”[All Fields])	11,280
17	Phyllodes tumour (“phyllodes tumour”[All Fields] OR “phyllodes tumor”[MeSH Terms] OR (“phyllodes”[All Fields] AND “tumor”[All Fields]) OR “phyllodes tumor”[All Fields])	2816
18	12 OR 13 OR 14 OR 15 OR 16 OR 17	585,714
19	Breast (“breast”[MeSH Terms] OR “breast”[All Fields] OR “breasts”[All Fields] OR “breast s”[All Fields])	646,508
20	Breast AND 18	85,995
21	breast cancer (“breast neoplasms”[MeSH Terms] OR (“breast”[All Fields] AND “neoplasms”[All Fields]) OR “breast neoplasms”[All Fields] OR (“breast”[All Fields] AND “cancer”[All Fields]) OR “breast cancer”[All Fields])	521,128
22	20 OR 21	583,145
23	Diagnosis, computer assisted (Mesh) (“diagnosis, computer assisted”[MeSH Terms] OR (“diagnosis”[All Fields] AND “computer assisted”[All Fields]) OR “computer-assisted diagnosis”[All Fields] OR (“diagnosis”[All Fields] AND “computer”[All Fields] AND “assisted”[All Fields]) OR “diagnosis computer assisted”[All Fields])	129,620
24	Prediction (“predict”[All Fields] OR “predictabilities”[All Fields] OR “predictability”[All Fields] OR “predictable”[All Fields] OR “predictably”[All Fields] OR “predicted”[All Fields] OR “predicting”[All Fields] OR “prediction”[All Fields] OR “predictions”[All Fields] OR “predictive”[All Fields] OR “predictively”[All Fields] OR “predictiveness”[All Fields] OR “predictives”[All Fields] OR “predictivities”[All Fields] OR “predictivity”[All Fields] OR “predicts”[All Fields])	1,997,298
25	23 OR 24	2,107,769
26	Screening (“diagnosis”[MeSH Subheading] OR “diagnosis”[All Fields] OR “screening”[All Fields] OR “mass screening”[MeSH Terms] OR (“mass”[All Fields] AND “screening”[All Fields]) OR “mass screening”[All Fields] OR “early detection of cancer”[MeSH Terms] OR (“early”[All Fields] AND “detection”[All Fields] AND “cancer”[All Fields]) OR “early detection of cancer”[All Fields] OR “screen”[All Fields] OR “screenings”[All Fields] OR “screened”[All Fields] OR “screens”[All Fields])	6,206,466
27	mammogram (“mammography”[MeSH Terms] OR “mammography”[All Fields] OR “mammogram”[All Fields] OR “mammograms”[All Fields])	45,728
28	Imaging (“image”[All Fields] OR “image s”[All Fields] OR “imaged”[All Fields] OR “imager”[All Fields] OR “imager s”[All Fields] OR “imagers”[All Fields] OR “images”[All Fields] OR “imaging”[All Fields] OR “imaging s”[All Fields] OR “imagings”[All Fields])	3,048,404
29	Image Interpretation, Computer-Assisted (“image interpretation, computer assisted”[MeSH Terms] OR (“image”[All Fields] AND “interpretation”[All Fields] AND “computer assisted”[All Fields]) OR “computer-assisted image interpretation”[All Fields] OR (“image”[All Fields] AND “interpretation”[All Fields] AND “computer”[All Fields] AND “assisted”[All Fields]) OR “image interpretation computer assisted”[All Fields])	618,833
30	Image Processing, Computer-Assisted (“image processing, computer assisted”[MeSH Terms] OR (“image”[All Fields] AND “processing”[All Fields] AND “computer assisted”[All Fields]) OR “computer-assisted image processing”[All Fields] OR (“image”[All Fields] AND “processing”[All Fields] AND “computer”[All Fields] AND “assisted”[All Fields]) OR “image processing computer assisted”[All Fields])	269,922

(continued on next page)

(*continued*)

#	Search Terms for Pubmed/Medline Version 2	Results
31	Magnetic Resonance Imaging OR MRI (“magnetic resonance imaging”[MeSH Terms] OR (“magnetic”[All Fields] AND “resonance”[All Fields] AND “imaging”[All Fields]) OR “magnetic resonance imaging”[All Fields] OR (“magnetic resonance imaging”[MeSH Terms] OR (“magnetic”[All Fields] AND “resonance”[All Fields] AND “imaging”[All Fields]) OR “magnetic resonance imaging”[All Fields] OR “mri”[All Fields])	782,520
32	Ultrasound (“diagnostic imaging”[MeSH Subheading] OR (“diagnostic”[All Fields] AND “imaging”[All Fields]) OR “diagnostic imaging”[All Fields] OR “ultrasound”[All Fields] OR “ultrasonography”[MeSH Terms] OR “ultrasonography”[All Fields] OR “ultrasonics”[MeSH Terms] OR “ultrasonics”[All Fields] OR “ultrasounds”[All Fields] OR “ultrasound s”[All Fields])	2,045,849
33	25 OR 26 OR 27 OR 28 OR 29 OR 30 OR 31 OR 32	8,843,740
34	5 AND 11 AND 22 AND 33	3434

Appendix 2. Data extraction form

Dataset name
Dataset URL
Dataset DOI (if available)
Dataset name
Is the documentation a single document / resource, or is it split across multiple sources?
What does this dataset's documentation consist of?
How is the dataset accessed?
What country/countries do the PATIENTS come from? Tick all that apply (please leave blank if insufficient information to answer this confidently).
Which country/countries do DATASET CURATORS come from? Tick all that apply (leave blank if insufficient data to answer confidently).
Are specific collection sites listed? If so, what are they? (if not reported, please state so).
Is the data derived from a screening programme or conducted as a diagnostic test?
If mammograms are labelled, how were they generated?
If mammograms are labelled, how were ground truths about diagnosis determined?
Is a scoring system included? If yes, please describe here.
Were the data derived from a private or public healthcare setting?
How many individuals are represented in this dataset?
Regarding consent and ethics

The dataset received ethical approval or waiver by an ethics committee, IRB or similar
Patients gave their consent for data to be shared

Any other comments?

1.1 Dataset Summary

Regarding the dataset documentation:

Describes the contents of the dataset
Describes the source of the dataset
Written in accessible language
Would help data users assess if the dataset meets their needs

Comments on item 1.1 (dataset summary)

1.2 Dataset identity and access

Dataset documentation should include: dataset name, accessibility, date of release, version, licensing arrangements, and details of the data custodian(s). Where possible this documentation should adhere to FAIR principles.

Regarding the dataset documentation:

Contains the dataset name
Contains details on how to access the dataset
Contains the date of release
Contains the version number
Contains details of licensing arrangements
Contains details of data custodian(s)
In what format are images stored?

Comments on item 1.2 (dataset identity and access)

1.3 Motivations for dataset creation and intended purpose(s)

Dataset documentation should include the reasons why this dataset was created, including any intended benefit(s), any purposes for which dataset use should be avoided, who created the dataset, and who funded it. Regarding the dataset documentation:

Includes reasons why dataset was created
Includes any intended benefits
Includes purposes for which dataset use should be avoided
States who created the dataset
States who funded the dataset

Comments on item 1.3 (motivations for dataset creation and intended purposes)

1.4 Assumptions and preconceptions of the dataset curation team

Dataset documentation should describe how the curation team has considered the impact of their prior assumptions and preconceptions on biases in the dataset. This may include reflecting on the experiences of the dataset curators themselves, as well as any advice from governing and consultation groups (e.g. advisory boards, patient and public involvement and engagement groups). Regarding the dataset documentation:

Describes how the curation team considered impact of prior assumptions and preconceptions on biases in the dataset
Describes experience of dataset curators themselves
Includes advice from advisory boards
Includes advice from PPIE groups
Includes advice from other groups

Comments on item 1.4 (assumptions and preconceptions of the dataset curation team)

1.5 Origin and purpose of source data

Dataset documentation should describe the original source of data, including the reason for generating the original data (e.g., patient records to provide clinical care, clinical trial, biobank) and what individuals were expecting to happen to their data (e.g., administrative action, participant in a research study). Regarding the dataset documentation:

Describes original source of data
Gives reason for generating original data
Describes what individuals were expecting to happen to their data

Comments on item 1.5 (origin and purpose of source data)

1.6 Data sampling, and aggregation from multiple sources

Dataset documentation should describe how data were sampled from the original data source, including an explanation of sampling strategies and their rationale. If the dataset has been compiled from multiple data sources, dataset documentation should describe how datasets were selected, and how decisions were made during data aggregation, particularly in the case of grouping populations and modification of demographic coding. Regarding the dataset documentation:

Describes how data were sampled
Includes explanation of sampling strategy
Includes rationale for sampling strategy
If multiple datasets combined, describes how datasets were selected
If multiple datasets combined, describes how decisions were made during data aggregation
If multiple datasets combined, comments on how demographic codes were grouped

Comments on item 1.6 (data sampling and aggregation from multiple sources)

1.7 Data shifts

For longitudinal datasets or datasets with versions, dataset documentation should describe any known or suspected changes over time relating to the population, medical practice, or how data were collected, which may contribute to data shifts. Regarding the dataset documentation:

Describe known or suspected changes over time which may contribute to data shifts

Comments on item 1.7 (data shifts)

1.8 Composition of populations

Dataset documentation should: - Include a summary of the populations present in the dataset. The choice of which populations to describe, and the choice of grouping/categorisation, should be explained. - Highlight any known missing groups within the dataset and any reason(s) for their missingness. Regarding the dataset documentation:

Includes a summary of populations present in the dataset
Explains choice of which populations are described
Explains population grouping/categorisation
Highlight known missing groups
Explain why groups are missing

Comments on item 1.8 (composition of populations)

1.9 Recording of attributes of individuals

Dataset documentation should: - Describe how and why attributes are provided in the dataset (self-reported by participants, imputed, linked from other datasets), and whether this information is available at the individual or aggregate level. - Explain whether attributes have been coded, condensed or modified, stating how and why this was done. - Highlight the proportion of attributes recorded as ‘unknown’ or ‘other’, and if possible explain the reasons why Regarding the dataset documentation:

Describes how and why attributes are provided
States whether information is provided at individual / aggregate level
Explains whether attributes have been coded/condensed/modified
States why attributes have been coded/condensed/modified
Highlights proportion of attributes recorded as ‘unknown’ or other
Explains why some attributes are recorded as ‘unknown’ or other

Comments on item 1.9 (recording of attributes of individuals)

1.10 Groups at particular risk of harm

Dataset documentation should: - Always include data (when available) on certain attributes (including age, gender identity, sex, race, ethnicity, socioeconomic status, and sexual orientation), due to known associations with health outcomes and interactions with wider social factors. If including these data may place individuals at risk of identification or endanger them, these data should instead be provided at aggregate level for the whole dataset. If data on these particular attributes are missing, reasons for this should be stated. - Highlight the presence of any vulnerable population groups in this dataset, with consideration of both vulnerabilities that are universal (e.g., children, people with severe disabilities, displaced persons) and those that are specific to the site of data collection (e.g., marginalised religious or caste groups). Regarding the dataset documentation:

Has data on age (or states not available)
Has data on gender identity (or states not available)
Has data on sex (or states not available)
Has data on race (or states not available)
Has data on ethnicity (or states not available)
Has data on socioeconomic status (or states not available)
Has data on sexual orientation (or states not available)
Highlights vulnerable population groups
If age is reported, please provide a summary (ideally by stating the median, interquartile range, and overall range)
Sometimes it isn't clear what is meant by sex/gender as labels. If sex/gender are included, please a) give the explanation of what they mean (if not provided, please state so) and b) describe how sex/gender are categorised and how many people are represented in each group
If race and/or ethnicity are reported, what categories are used? What definition of race/ethnicity do they give? How many people are represented within each group?

Comments on item 1.10 (groups at particular risk of harm)

1.11 Modifications made to the data

Dataset documentation should describe whether any data items were modified from the original source, providing the rationale for doing so and any methods used. For example; for anonymisation, to correct for imbalance, to correct errors or biases, or to enable mapping to existing data standards. Regarding the dataset documentation:

Describes whether any data items were modified from the original source
Provides rationale for modification
Explains methods used for modification
Comments on item 1.11 (Modifications made to the data)

1.12 Limitations of the dataset

Dataset documentation should identify known or suspected sources of bias, error or other factors that affect the dataset as a whole, which may impact its generalisability or applicability for other use. Regarding the dataset documentation:

Identifies known or suspected sources of bias
Identifies known or suspected sources of error
Identifies other factors which affect dataset as a whole which may impact its generalisability or applicability

Comments on item 1.12 (limitations of the dataset)

1.13 Missing data

Dataset documentation should describe the proportion, nature and causes of missing data (if known), particularly if there are systematic differences across relevant population groups. Documentation should also describe if missing data have been identified and how they have been handled (e.g. imputation, correction). Regarding the dataset documentation:

Describes proportion of missing data
Describes nature of missing data
Describes causes of missing data
Describes if missing data is different or the same across relevant population groups
Describes how missing data have been handled

Comments on item 1.13 (missing data)

1.14 Errors in the data

Dataset documentation should: - Describe how errors can be/have been identified in the data and how they have been handled (e.g. have they been removed, modified, corrected or left in the dataset). - Provide an estimation of the proportion of errors that are present and whether they are more prevalent in some population groups than others. - Provide possible reasons for any systematic differences in error rates across population groups within the dataset. Regarding the dataset documentation:

- Describes how errors can be or have been identified.
- Describes how errors have been handled.
- Errors have been removed.
- Errors have been corrected.
- Errors have been modified.
- Errors have been left in the dataset.
- Describes the proportion of errors that are present.
- Describes whether errors are more prevalent in some population groups than others
- Gives possible reasons for differences in error rates across population groups

Comments on item 1.14 (errors in the data)

1.15 Known or potential bias in data generation

Dataset documentation should: - Describe how bias may be introduced by the acquisition and processing of data within the dataset. For example: from the use of devices, sensors and software. - Highlight any known or potential differences in data acquired across different population groups, or any uncertainty of measurements within population groups. - Describe any attempts to mitigate these biases. Regarding the dataset documentation:

Describes how acquisition of data may have introduced bias
Describes how processing of data may have introduced bias
Highlights known or potential differences in data acquired across population groups
Highlights uncertainty of measurements within population groups
Describes attempts to mitigate these biases

Comments on item 1.15 (known or potential bias in data generation)

1.16 Known or potential bias in data collection

Dataset documentation should: - Identify areas where bias may have been introduced into the data collection process. For example: only collecting data from one geographical area, context regarding healthcare coverage and accessibility, only using questionnaires in English. - Describe any attempts to mitigate these biases. Regarding the dataset documentation:

Identifies areas where bias may have been introduced into the data collection process
Describes attempts to mitigate these biases

Comments on item 1.16 (known or potential bias in data collection)

1.17 Known or potential bias in data labels

Dataset documentation should: - Provide a description of any data labels, including who decided what labels to include, what they were called, and how they were generated. - Highlight labels that are at high risk of bias. For example, where label generation was at the discretion of individuals, where known biases in labelling behaviour has been evidenced previously, or in the use of proxy variables (e.g., healthcare costs as a proxy of healthcare needs) - Describe any attempts to mitigate these biases. Regarding the dataset documentation:

Provides a description of data labels
States who decided what labels to include
States who decided what labels should be called
States how labels were generated
Highlights labels at high risk of bias
Describes attempts to mitigate these biases
Describes attempts to mitigate these biases

Comments on item 1.17 (known or potential bias in data labels)

1.18 Ethics, governance, and quality assurance

Dataset curators should state in their documentation whether data protection laws specific to their jurisdiction have been adhered to. Dataset documentation should also: - Describe measures taken to protect the identities of individuals. - Describe permissions obtained to enable dataset curation, and details of the governance of the dataset. - Provide references to institutional review board/ethical committee review (or equivalent, as appropriate). - Reference standards (e.g. ISO, FAIR) which have been adhered to. Regarding the dataset documentation:

States whether data protection laws have been adhered to
Describes measures taken to protect the identity of individuals
Describes permissions obtained to enable dataset curation
Describes governance of the dataset
Provides references to IRB/ethics committee review (or equivalent)
References standards adhered to (e.g. ISO, FAIR)

Comments on item 1.18 (ethics, governance, and quality assurance)

1.19 Patient and public involvement and engagement

Dataset documentation should: - Describe the role of any advisory boards and patient and public involvement and engagement groups in the dataset curation. - Provide information on any efforts to share data and findings with those who contributed to the dataset and any feedback gathered from participants that is relevant to data interpretation. Regarding the dataset documentation:

Describes the role of advisory boards
Describes the role of patient and public involvement and engagement groups
Provides information on efforts to share data/findings with those who contributed data to the dataset
Describes feedback from participants that is relevant to data interpretation

Comments on item 1.19 (patient and public involvement and engagement)

1.20 Bias and impact assessments

If a formal assessment of bias, fairness or societal impact has been previously conducted on the dataset, dataset documentation should provide these assessments and results. This may include algorithmic impact assessments (AIAs), data protection impact assessments (DPIAs), equality impact assessments, documentation tools, risk of bias assessments or automated toolkits. Regarding the dataset documentation:

Provides details of any assessment of bias, fairness or societal impact
Provides details of algorithmic impact assessment (AIA)
Provides details of data protection impact assessment (DPIA)
Provides details of equality impact assessment
Provides details of risk of bias assessments, documentation tools or automated toolkits

Comments on item 1.20 (Bias and impact assessment

Appendix 3. Datasets that we were unable to access

*Where authors discussed several datasets used, we analysed each separately for country of origin, number of individuals, screening/diagnostic and whether the original mammograms were digitised from film images.

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
A Multiple Circular Path Convolution Neural Network System for Detection of Mammographic Masses	Lo SC, Li H, Wang Y, Kinnard L and Freedman MT	https://pubmed.ncbi.nlm.nih.gov/11929102/	2002	Brook Army Medical Center Database	Not reported	Not reported	Not reported	Not reported	Not reported
A Multi-million Mammography Image Dataset and Population-Based Screening Cohort for the Training and Evaluation of Deep Neural Networks—the Cohort of Screen-Aged Women (CSAW)	Dembrower K, Lindholm P and Strand F	https://pubmed.ncbi.nlm.nih.gov/31520277/	2020	Cohort of Screen Aged Women	Sweden	499,807	1,182,733	Screening	Not reported
Convolutional neural network for breast cancer diagnosis using diffuse optical tomography	Xu Q and Wang X and Jiang H	https://pubmed.ncbi.nlm.nih.gov/32240400/	2019		China	Not reported	63	Not reported	Not reported
Application of deep learning in the detection of breast lesions with four different breast densities	Li H, Ye J, Liu H, Wang Y, Shi B, Chen J, Kong A, Xu Q and Cai J	https://pubmed.ncbi.nlm.nih.gov/34132495/	2021		China	58,516	Not reported	Not reported	Not reported
*Deep-LIBRA: An artificial-intelligence method for robust quantification of breast density with independent validation in breast cancer risk assessment	Haji Maghsoudi O, Gastounioti A, Scott C, Pantalone L, Wu FF, Cohen EA, Winham S, Conant EF, Vachon C and Kontos D	https://pubmed.ncbi.nlm.nih.gov/34274690/	2021	Maghsoudi et al. ds1 Maghsoudi et al. ds3a Maghsoudi et al. ds3b Maghsoudi et al. ds5	USA USA USA USA	2200 1662 575 1592	11,200 3314 1147 6368	Screening Not reported Not reported	Not reported Not reported Not reported
Classification of microcalcification clusters in digital breast tomosynthesis using ensemble convolutional neural network	Xiao B, Sun H, Meng Y, Peng Y, Yang X, Chen S, Yan Z and Zheng J	https://pubmed.ncbi.nlm.nih.gov/34320986/	2021		China	236	Not reported	Not reported	Not reported
Connected-UNets: a deep learning architecture for breast mass segmentation	Baccouche A, Garcia-Zapirain B, Olea CC and Elmaghraby AS	https://pubmed.ncbi.nlm.nih.gov/34857755/	2021		Mexico	208	Not reported	Not reported	Not reported
Improving the classification ability of network utilizing fusion technique in contrast-enhanced spectral mammography	Song J, Zheng Y, Xu C, Zou Z, Ding G and Huang W	https://pubmed.ncbi.nlm.nih.gov/34860417/	2022		China	95	Not reported	Not reported	Not reported
Mammographically occult breast cancers detected with AI-based diagnosis supporting software: clinical and histopathologic characteristics	Kim HJ, Kim HH, Kim KH, Choi WJ, Chae EY, Shin HJ, Cha JH and Shim WH	https://pubmed.ncbi.nlm.nih.gov/35347508/	2022		South Korea	1762	Not reported	Both	Not reported
Improving the Performance of Radiologists Using Artificial Intelligence-Based Detection Support Software for Mammography: A Multi-Reader Study	Lee JH, Kim KH, Lee EH, Ahn JS, Ryu JK, Park YM, Shin GW, Kim YJ and Choi HY	https://pubmed.ncbi.nlm.nih.gov/35434976/	2022		South Korea	200	Not reported	Screening	Not reported

(continued on next page)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
An automated confirmatory system for analysis of mammograms	Peng W, Mayorga RV and Hussein EM	https://pubmed.ncbi.nlm.nih.gov/26742491/	2015	BancoWeb Database	Brazil	Not reported	1704	Not reported	Not reported
Evaluation of deep learning-based artificial intelligence techniques for breast cancer detection on mammograms: Results from a retrospective study using a BreastScreenVictoria dataset	Frazer HM, Qin AK, Pan H and Brotchie P	https://pubmed.ncbi.nlm.nih.gov/34212526/	2021	BreastScreen Victoria dataset	Australia	7498	28,694	Screening	No
Multi- class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN)	Heenaye-Mamode Khan M, Boodoo-Jahangeer N, Dullull W, Nathire S, Gao X, Sinha GR and Nagwanshi KK	https://pubmed.ncbi.nlm.nih.gov/34437623/	2021	UPMC	Not reported	Not reported	Not reported	Not reported	Not reported
Artificial Intelligence Detection of Missed Cancers at Digital Mammography That Were Detected at Digital Breast Tomosynthesis	Dahlblom V, Andersson I, Lång K, Tingberg A, Zackrisson S and Dustler M	https://pubmed.ncbi.nlm.nih.gov/34870215/	2021	Malmö Breast Tomosynthesis Screening Trial (MBTST)	Sweden	14,768	Not reported	Screening	Not reported
Genetic programming and feature selection for classification of breast masses in mammograms	Nandi RJ, Nandi AK, Rangayyan R and Scutt D	https://pubmed.ncbi.nlm.nih.gov/17945751/	2006	Alberta Cancer Board, "Screen Test: Alberta Program for the Early Detection of Breast Cancer"	Canada	Not reported	57	Screening	Not reported
SVM and SVM Ensembles in Breast Cancer Prediction	Huang MW, Chen CW, Lin WC, Ke SW and Tsai CF	https://pubmed.ncbi.nlm.nih.gov/28060807/	2017	SIGKDD Cup 2008 http://www.sigkdd.org/kddcup/index.php	Not reported	Not reported	102,294	Not reported	Not reported
Detecting and classifying lesions in mammograms with Deep Learning	Ribli D, HorvÃ¡th A, Unger Z, Pollner P and Csabai I	https://pubmed.ncbi.nlm.nih.gov/29545529/	2018	Semmelweis University Budapest dataset	Hungary	174	847	Not reported	Not reported
Local Binary Patterns Descriptor Based on Sparse Curvelet Coefficients for False-Positive Reduction in Mammograms	Pawar MM, Talbar SN and Dudhane A	https://pubmed.ncbi.nlm.nih.gov/30356422/	2018		India	90	360	Not reported	Not reported
Development and validation of a deep learning model for detection of breast cancers in mammography from multi-institutional datasets	Ueda D, Yamamoto A, Onoda N, Takashima T, Noda S, Kashiwagi S, Morisaki T, Fukumoto S, Shiba M, Morimura M, Shimono T, Kageyama K, Tatekawa H, Murai K, Honjo T, Shimazaki A, Kabata D and Miki Y	https://pubmed.ncbi.nlm.nih.gov/35324962/	2022	Hospital development dataset	Japan	Not reported	3179	Not reported	Not reported
Diagnosis of Breast Cancer Using Radiomics Models Built Based on Dynamic Contrast Enhanced MRI	Zhao YF, Chen Z, Zhang Y, Zhou J, Chen JH, Lee KE, Combs FJ, Parajuli R, Mehta RS, Wang M and Su MY	https://pubmed.ncbi.nlm.nih.gov/34869020/	2021		China	Not reported	Not reported	Diagnostic	No

(continued on next page)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
Combined With Mammography									
A real use case of semi-supervised learning for mammogram classification in a local clinic of Costa Rica	Calderon-Ramirez S, Murillo-Hernandez D, Rojas-Salazar K, Elizondo D, Yang S, Moemeni A, and Molina-Cabello M	https://pubmed.ncbi.nlm.nih.gov/35239108/	2022	Clínica Chavarría's 2020 mammogram target dataset	Costa Rica	87	341	Not reported	Not reported
Classification of clustered microcalcifications using a Shape Cognitron neural network	Lee SK, Chung PC, Chang CI, Lo CS, Lee T, Hsu GC and Yang CW	https://pubmed.ncbi.nlm.nih.gov/12576111/	2003	Nijmegen mammogram database	Netherlands	Not reported	40	Not reported	Not reported
Computer-aided mass detection on digitized mammograms using adaptive thresholding and fuzzy entropy	Younesi F, Alam N, Zoroofi RA, Ahmadian A and Guiti M	https://pubmed.ncbi.nlm.nih.gov/18003291/	2007	BIRADS	Not reported	58		Not reported	Not reported
A new method of detecting micro-calcification clusters in mammograms using contourlet transform and non-linking simplified PCNN	Guo Y, Dong M, Yang Z, Gao X, Keju Wang K, Luo C, Ma Y, Zhang J	https://pubmed.ncbi.nlm.nih.gov/27208519/	2016	JSMIT	Japan	Not reported	11	Not reported	Yes
Deep Learning Pre-training Strategy for Mammogram Image Classification: an Evaluation Study	Kadie Clancy, Sarah Aboutalib, Aly Mohamed, Jules Sumkin, Shandong Wu	https://pubmed.ncbi.nlm.nih.gov/32607908/	2020	FFDM dataset	USA	1303	4935	Screening	No
Presentation of Novel Hybrid Algorithm for Detection and Classification of Breast Cancer Using Growth Region Method and Probabilistic Neural Network	Isfahani ZN, Jannat-Dastjerdi I, Eskandari F, Ghoushchi SJ and Pourasad Y	https://pubmed.ncbi.nlm.nih.gov/34239550/	2021	BI-RADS	Not reported	Not reported	60	Screening	Not reported
A Deep Learning Approach to *Re*-create Raw Full-Field Digital Mammograms for Breast Density and Texture Analysis	Shu H, Chiang T, Wei P, Do KA, Lesslie MD, Cohen EO, Srinivasan A, Moseley TW, Chang Sen LQ, Leung JWT, Dennison JB, Hanash SM and Weaver OO	https://pubmed.ncbi.nlm.nih.gov/34350403/	2021	MERIT study	USA	884	Not reported	Screening	Not reported
Artificial Intelligence for Breast Cancer Screening in Mammography (AI-STREAM): A Prospective Multicenter Study Design in Korea Using AI-Based CADe/x	Chang YW, An JK, Choi N, Ko KH, Kim KH, Han K, Ryu JK	https://pubmed.ncbi.nlm.nih.gov/35133093/	2022	National cancer registry database	South Korea	32,714	Not reported	Screening	Not reported
*Development and validation of a deep learning model for detection of breast cancers in mammography from multi-institutional datasets	Ueda D, Yamamoto A, Onoda N, Takashima T, Noda S, Kashiwagi S, Morisaki T, Fukumoto S, Shiba M, Morimura M, Shimono T, Kageyama K, Tatekawa H, Murai K, Honjo T,	https://pubmed.ncbi.nlm.nih.gov/35324962/	2022	Hospital test dataset Clinic test dataset	Japan Japan	Not reported Not reported	491 2821	Not reported Screening	Not reported Not reported

(*continued on next page*)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
	Shimazaki A, Kabata D and Miki Y								
Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network	Wang XH, Zheng B, Good WF, King JL and Chang YH	https://pubmed.ncbi.nlm.nih.gov/10219951/	1999		USA	419	Not reported	Not reported	Yes
Computer-aided, Case-based Diagnosis of Mammographic Regions of Interest Containing Microcalcifications	Jack Sklansky, Eric Y. Tao, Mohsen Bazargan, Chester J. Ornes, Robert C. Murchison,Senait Teklehaimanot	https://pubmed.ncbi.nlm.nih.gov/10845398/	2000		USA	138	Not reported	Not reported	Yes
Computerized Classification of Benign and Malignant Masses on Digitized Mammograms: A Study of Robustness	Zhimin Huo, Maryellen L. Giger, Carl J. Vyborny, Dulcy E. Wolverton, Charles E. Metz	https://pubmed.ncbi.nlm.nih.gov/11131052/	2000		USA	175	307	Not reported	Yes
Dependence of computer classification of clustered microcalcifications on the correct detection of microcalcifications	Jiang Y, Nishikawa RM and Papaioannou J	https://pubmed.ncbi.nlm.nih.gov/11585226/	2001		Not reported	53	100	Not reported	Yes
Optimal Neural Network Architecture Selection: Improvement in Computerized Detection of Microcalcifications	Gurcan MN, Chan HP, Sahiner B, Hadjiiski L, Petrick N and Helvie MA	https://pubmed.ncbi.nlm.nih.gov/11942656/	2002		USA	Not reported	260	Both	No
Advances in micro-calcification clusters detection in mammography	Zhang L, Sankar R and Qian W	https://pubmed.ncbi.nlm.nih.gov/12356499/	2002	USUHS	USA	67	97	Not reported	No
Computer-aided diagnosis of mammographic microcalcification clusters	Kallergi M	https://pubmed.ncbi.nlm.nih.gov/15000617/	2004		USA	Not reported	100	Not reported	Yes
A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography	Sheila Timp, Nico Karssemeijer	https://pubmed.ncbi.nlm.nih.gov/15191279/	2004		Netherlands	Not reported	4295	Screening	Yes
Classification of breast masses in mammograms using genetic programming and feature selection	Nandi RJ, Nandi AK, Rangayyan RM and Scutt D	https://pubmed.ncbi.nlm.nih.gov/16937210/	2006		USA	Not reported	57	Screening	Yes
A completely automated CAD system for mass detection in a large mammographic database	R Bellotti, F De Carlo, S Tangaro, G Gargano, G Maggipinto, M Castellano, R Massafra, D Cascio, F Fauci, R Magro, G Raso, A Lauria, G Forni, S Bagnasco, P Cerello, E Zanon, S C Cheran, E Lopez Torres, U Bottigli, G L Masala, P Oliva, A Retico, M E	https://pubmed.ncbi.nlm.nih.gov/16964885/	2006		Italy	967	3369	Both	Yes

(continued on next page)

(continued)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
	Fantacci, R Cataldo, I De Mitri, G De Nunzio								
Dual system approach to computer-aided detection of breast masses on mammograms	Wei J and Chan HP and Sahiner B and Hadjiiski LM and Helvie MA and Roubidoux MA and Zhou C and Ge J	https://pubmed.ncbi.nlm.nih.gov/17153394/	2009		USA	180	Not reported	Not reported	Yes
Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier	Delogua P, Fantaccia ME, Kasaeb P, Retico A	https://pubmed.ncbi.nlm.nih.gov/17383623/	2019		Italy	Not reported	Not reported	Both	Yes
A new parameter enhancing breast cancer detection in computer-aided diagnosis of x-ray mammograms	Tanki N, Murase K and Nagao M	https://pubmed.ncbi.nlm.nih.gov/17634739/	2006		Japan	60	Not reported	Not reported	Yes
Temporal change analysis for characterization of mass lesions in mammography	Timp S, Varela C and Karssemeijer N	https://pubmed.ncbi.nlm.nih.gov/17649908/	2007		Netherlands	Not reported	465	Screening	Yes
An ellipse-fitting based method for efficient registration of breast masses on two mammographic views	Pu J, Zheng B, Leader JK and Gur D	https://pubmed.ncbi.nlm.nih.gov/18383669/	2008		Not reported	Not reported	400	Not reported	Yes
GPCALMA: implementation in Italian hospitals of a computer aided detection system for breast lesions by mammography examination	Lauria A	https://pubmed.ncbi.nlm.nih.gov/18602854/	2009	GPCALMA	Italy	967	3369	Screening	Yes
Markov random field-based clustering applied to the segmentation of masses in digital mammograms	Suliga M, Deklerck R and Nyssen E	https://pubmed.ncbi.nlm.nih.gov/18620842/	2008		Not reported	Not reported	100	Not reported	Not reported
Computer-aided detection of masses in full-field digital mammography using screen-film mammograms for training	Michiel Kallenberg, Nico Karssemeijer	https://pubmed.ncbi.nlm.nih.gov/19001703/	2008		Netherlands	266	Not reported	Not reported	Yes
Correlative feature analysis on FFDM	Yuan Y, Giger ML, Li H and Sennett C	https://pubmed.ncbi.nlm.nih.gov/19175108/	2008		USA	Not reported	Not reported	Not reported	No
A review of automatic mass detection and segmentation in mammographic images	Arnau Oliver, Jordi Freixenet, Joan Martí, Elsa Pérez, Josep Pont, Erika R E Denton, Reyer Zwiggelaar	https://pubmed.ncbi.nlm.nih.gov/20071209/	2010		Spain	Not reported	176	Not reported	Not reported
Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis	Wei L, Yang Y and Nishikawa RM	https://pubmed.ncbi.nlm.nih.gov/20161326/	2009		USA	104	200	Not reported	Yes
Influence of Using Manual or Automatic Breast Density	Oliver A, Llado X, Freixenet J, Marta	https://pubmed.ncbi.nlm.	2010		Not reported	Not reported	184	Not reported	No

(continued on next page)

(continued)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
Information in a Mass Detection CAD System	R, Perez E, Pont J and Zwiggelaar R	nih.gov/20540910/							
Breast cancer risk estimation with artificial neural networks revisited: discrimination and calibration	Turgay Ayer, Oguzhan Alagoz, Jagpreet Chhatwal, Jude W Shavlik, Charles E Kahn Jr., Elizabeth S Burnside	https://pubmed.ncbi.nlm.nih.gov/20564067/	2010		USA	18,269	48,744	Both	No
Characterization of masses in digital breast tomosynthesis: comparison of machine learning in projection views and reconstructed slices	Heang-Ping Chan, Yi-Ta Wu, Berkman Sahiner, Jun Wei, Mark A Helvie, Yiheng Zhang, Richard H Moore, Daniel B Kopans, Lubomir Hadjiiski, Ted Way	https://pubmed.ncbi.nlm.nih.gov/20831065/	2010		USA	99	107	Not reported	No
Detection of clustered microcalcifications using spatial point process modeling	Jing H, Yang Y and Nishikawa RM	https://pubmed.ncbi.nlm.nih.gov/21119233/	2011		USA	66	141	Not reported	Yes
Image feature evaluation two new mammography CAD prototypes	Hapfelmeier A and Horsch A	https://pubmed.ncbi.nlm.nih.gov/21380554/	2011		Germany	306	1027	Screening	No
Mammographic masses characterization based on localized texture and dataset fractal analysis usingl inear, neural and support vector machine classifiers	Murase K, Tanki N, Miyazaki S and Nagao M	https://pubmed.ncbi.nlm.nih.gov/21976251/	2008		Not reported	Not reported	130	Not reported	Yes
*Automatic Detection of Pectoral Muscle Using Average Gradient and Shape Based Feature	Chakraborty J, Mukhopadhyay S, Singla V, Khandelwal N, and Bhattacharyya P	https://pubmed.ncbi.nlm.nih.gov/22006275/	2012	Direct radiography images Computed radiography images	Not reported Not reported	Not reported Not reported	80 40	Not reported Not reported	Yes Not reported
Measures of angular spread and entropy for the detection of architectural distortion in prior mammograms	Banik S, Rangayyan RM, Desautels JEL	https://pubmed.ncbi.nlm.nih.gov/22460365/	2012		Canada	Not reported	158	Screening	Yes
Measures of divergence of oriented patterns for the detection of architectural distortion in prior mammograms	Rangayyan RM, Banik S, Chakraborty J, Mukhopadhyay S and Desautels JE	https://pubmed.ncbi.nlm.nih.gov/23054747/	2012		Canada	69	210	Screening	Yes
Breast mass contour segmentation algorithm in digital mammograms	Berber T, Alpkocak A, Balci P, Dicle O	https://pubmed.ncbi.nlm.nih.gov/23273502/	2012	DEMS Dataset	Turkey	485	Not reported	Not reported	Not reported
Mass detection in reconstructed digital breast tomosynthesis volumes with a computer-aided detection system trained on 2D mammograms	van Schie G, Wallis MG, Leifland K, Danielsson M and Karssemeijer N	https://pubmed.ncbi.nlm.nih.gov/23556896/	2013		Not reported	220	Not reported	Both	No
Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector	Tan M, Pu J and Zheng B	https://pubmed.ncbi.nlm.nih.gov/24664267/	2015		Not reported	Not reported	Not reported	Not reported	Yes

(continued on next page)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
machine (SVM) model									
Reduction of false-positive recalls using a computerized mammographic image feature analysis scheme	Tan M, Pu J and Zheng B	https://pubmed.ncbi.nlm.nih.gov/25029964/	2014		Not reported	1052	Not reported	Both	No
A new and fast image feature selection method for developing an optimal mammographic mass detection scheme	Tan M, Pu J and Zheng B	https://pubmed.ncbi.nlm.nih.gov/25086537/	2014		Not reported	Not reported	Not reported	Not reported	Yes
A new approach to develop computer-aided detection schemes of digital mammograms	Tan M, Qian W, Pu J, Liu H, Zheng B	https://pubmed.ncbi.nlm.nih.gov/25984710/	2015		USA	1896	Not reported	Screening	No
Characterization of Architectural Distortionin Mammograms Basedon Texture Analysis Using Support Vector Machine Classifier with Clinical Evaluation	Kamra A, Jain VK, Singh S and Mittal S	https://pubmed.ncbi.nlm.nih.gov/26138756/	2016	ACE dataset	India	Not reported	Not reported	Not reported	Yes
Detection and classification of masses in mammographic images in a multi-kernel approach	Lima SML, Silva-Filho AG, Dos Santos WP	https://pubmed.ncbi.nlm.nih.gov/27480729/	2016		Not reported	Not reported	Not reported	Not reported	Not reported
Computerized breast cancer analysis system using three stage semi-supervised learning method	Sun W, Tseng TL, Zhang J and Qian W	https://pubmed.ncbi.nlm.nih.gov/27586481/	2016		Not reported	400	Not reported	Both	Not reported
Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography	Samala RK, Chan HP, Hadjiiski L, Helvie MA, Wei J and Cha K	https://pubmed.ncbi.nlm.nih.gov/27908154/	2016		USA	Not reported	2282	Not reported	No
Computer-aided classification of mammographic masses using visually sensitive image features	Wang Y, Aghaei F, Zarafshani A, Qiu Y, Qian W and Zheng B	https://pubmed.ncbi.nlm.nih.gov/27911353/	2017		Not reported	301	Not reported	Not reported	No
Improving computer-aided detection assistance in breast cancer screening by removal of obviously false-positive findings	Mordang JJ, Gubern-Macrida A, Bria A, Tortorella F, den Heeten G and Karssemeijer N	https://pubmed.ncbi.nlm.nih.gov/28182277/	2017		Netherlands	1837	Not reported	Screening	No
Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer	Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A	https://pubmed.ncbi.nlm.nih.gov/28212138/	2017		Switzerland	Not reported	3228	Not reported	No
A new approach to develop computer-aided diagnosis scheme of breast	Qiu Y, Yan S, Gundreddy RR, Wang Y, Cheng S, Liu H and Zheng B	https://pubmed.ncbi.nlm.	2017		Not reported	Not reported	Not reported	Not reported	No

(continued on next page)

(continued)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
mass classification using deep learning technology		nih.gov/28436410/							
Malignancy Detection on Mammography Using Dual Deep Convolutional Neural Networks and Genetically Discovered False Color Input Enhancement	Teare P, Fishman M, Benzaquen O, Toledano E, Elnekave E	https://pubmed.ncbi.nlm.nih.gov/28656455/	2017	Zebra Mammography Dataset (ZMDS)	Not reported	Not reported	1739	Not reported	Not reported
A novel and reliable computational intelligence system for breast cancer detection	Shirazi AZ, Javad S, Chabok SM, Mohammadi Z	https://pubmed.ncbi.nlm.nih.gov/28891042/	2018		Not reported	822	Not reported	Not reported	No
Applying Data-driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study	Kim EK, Kim HE, Han K, Kang BJ, Sohn YM, Woo OH and Lee CW	https://pubmed.ncbi.nlm.nih.gov/29426948/	2018		South Korea	Not reported	29,107	Both	No
Classification of Breast Masses Using a Computer-Aided Diagnosis Scheme of Contrast Enhanced Digital Mammograms	Danala G, Patel B, Aghaei F, Heidari M, Li J, Wu T, Zheng B	https://pubmed.ncbi.nlm.nih.gov/29748869/	2018		USA	111	Not reported	Not reported	No
Radiomics based detection and characterization of suspicious lesions on full field digital mammograms	Sapate SG, Mahajan A, Talbar SN, Sable N, Desai S and Thakur M	https://pubmed.ncbi.nlm.nih.gov/30119844/	2018		India	90	360	Screening	No
Determination of mammographic breast density using a deep convolutional neural network	Ciritsis A, Rossi C, De Martini IV, Eberhard M, Marcon M, Becker AS, Berger N, Boss A	https://pubmed.ncbi.nlm.nih.gov/30209957/	2018		Switzerland	5221	20,578	Diagnostic	No
Deep Learning to Distinguish Recalled but Benign Mammography Images in Breast Cancer Screening	Aboutalib SS, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, and Wu S	https://pubmed.ncbi.nlm.nih.gov/30309858/	2018	Full-field Digital Mammography (FFDM) Dataset	Not reported	1303	5212	Screening	No
Automatic Labeling of Special Diagnostic Mammography Views from Images and DICOM Headers	Lituiev DS, Trivedi H, Panahiazar M, Norgeot B, Seo Y, Franc B, Harnish R, Kawczynski M and Hadley D	https://pubmed.ncbi.nlm.nih.gov/30465142/	2018		Not reported	Not reported	4000	Not reported	No
*Automated pectoral muscle identification on MLO-view mammograms: Comparison of deep neural network to conventional computer vision	Ma X, Wei J, Zhou C, Helvie MA, Chan HP, Hadjiiski LM and Lu Y	https://pubmed.ncbi.nlm.nih.gov/30771257/	2019	Digitised film mammograms Digital mammograms	USA Not reported	Not reported 92	637 Not reported	Not reported Screening	Yes No
Breast mass detection and diagnosis using fused features with density	Wanga Z, Huang Y, Li M, Zhang H, Li C, Xinb J, and Qian W	https://pubmed.ncbi.nlm.nih.gov/30856154/	2019		China	120	480	Not reported	No
Preference-Sensitive Management of Post-Mammography Decisions in Breast Cancer Diagnosis	Ayvaci MUS, Alagoz O, Ahsen ME and Burnside ES	https://pubmed.ncbi.nlm.nih.gov/31031555/	2019		Not reported	18,270	65,892	Not reported	Yes
Predicting Breast Cancer by Applying Deep Learning to	Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A,	https://pubmed.ncbi.nlm.	2019		Israel	13,214	Not reported	Not reported	No

(continued on next page)

(continued)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
Linked Health Records and Mammograms	Melamed R, Barkan E, Herzel E, Naor S, Karavani E, Koren G, Goldschmidt Y, Shalev V, Rosen-Zvi M, Guindy M	nih.gov/31210611/							
A Deep Learning Model to Triage Screening Mammograms: A Simulation Study	Yala A, Schuster T, Miles R, Barzilay R and Lehman C	https://pubmed.ncbi.nlm.nih.gov/31385754/	2019		Not reported	80,818	Not reported	Screening	No
Applying a new quantitative image analysis scheme based on global mammographic features to assist diagnosis of breast cancer	Chen X, Zargari A, Hollingsworth AB, Liu H, Zheng B, Qiu Y	https://pubmed.ncbi.nlm.nih.gov/31443864/	2019		Not reported	275	Not reported	Screening	No
Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning using Deep Neural Nets	Samala RK, Heang-Ping Chan, Hadjiiski L, Helvie MA, Richter CD and Cha KH	https://pubmed.ncbi.nlm.nih.gov/31622238/	2020		USA	Not reported	324	Not reported	No
Deep learning modelling using normal mammograms for predicting breast cancer risk	Arefan D, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S	https://pubmed.ncbi.nlm.nih.gov/31667873/	2020		USA	226	Not reported	Screening	No
A similarity measure method fusing deep feature for mammogram retrieval	Wang Z, Xin J, Huang Y, Xu L, Ren J, Zhang H, Qian W, Zhang X, Liu J	https://pubmed.ncbi.nlm.nih.gov/31868727/	2020		China	Not reported	740	Not reported	No
International evaluation of an AI system for breast cancer screening	McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, Chesus M, Corrado GS, Darzi A, Etemadi M, Garcia-Vicente F, Gilbert FJ, Halling-Brown M, Hassabis D, Jansen S, Karthikesalingam A, Kelly CJ, King D, Ledsam JR, Melnick D, Mostofi H, Peng L, Reicher JJ, Romera-Paredes B, Sidebottom R, Suleyman M, Tse D, Young KC, De Fauw J and Shetty S	https://pubmed.ncbi.nlm.nih.gov/31894144/	2020		USA	2296	Not reported	Screening	Not reported
Breast microcalcifications detection based on fusing features with DTCWT	Wang Z, Xin J, Zhang Q, Gao S, Ma C, Ren J, Zhang H, Qian W, Zhu W, Zhang X and Liu J	https://pubmed.ncbi.nlm.nih.gov/31985483/	2020		China	100	100	Not reported	No
Multi-View Mammographic Density Classification by Dilated and Attention-Guided Residual Learning	Cheng Li, Jingxu Xu, Qiegen Liu, Yongjin Zhou, Lisha Mou, Zuhui Pu, Yong Xia, Hairong Zheng, Shanshan Wang	https://pubmed.ncbi.nlm.nih.gov/32012021/	2021		Not reported	500	1985	Not reported	No

(continued on next page)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
Inconsistent Performance of Deep Learning Models on Mammogram Classification	Wang X, Liang G, Zhang Y, Blanton H, Bessinger Z and Jacobs N	https://pubmed.ncbi.nlm.nih.gov/32068005/	2020	UKy	USA	507	1872	Not reported	No
Generalization error analysis for deep convolutional neural network with transfer learning in breast cancer diagnosis	Samala RK, Chan HP, Hadjiiski LM, Helvie MA and Richter CD	https://pubmed.ncbi.nlm.nih.gov/32208369/	2020		USA	Not reported	3411	Not reported	Yes
Deep learning of mammary gland distribution for architectural distortion detection in digital breast tomosynthesis	Li Y, He Z, Lu Y, Ma X, Guo Y, Xie Z, Qin G, Xu W, Xu Z, Chen W and Chen H	https://pubmed.ncbi.nlm.nih.gov/32485700/	2021		China	68	Not reported	Not reported	No
Prospective Analysis Usinga Novel CNN Algorithm to Distinguish Atypical Ductal Hyperplasia From Ductal Carcinoma in Situ in Breast	Mutasa S, Chang P, Nemer J, Van Sant EP, Sun M, McIlvride A, Siddique M and Ha R	https://pubmed.ncbi.nlm.nih.gov/32680766/	2020		Not reported	140	280	Not reported	No
Observational Study to Evaluate the Clinical Efficacy of Thermalytix for Detecting Breast Cancer in Symptomatic and Asymptomatic Women	Kakileti ST, Madhu HJ, Krishnan L, Manjunath G, Sampangi S and Ramprakash HV	https://pubmed.ncbi.nlm.nih.gov/33001739/	2020		India	470	Not reported	Both	No
Visualizing “featureless” regions on mammograms classified as invasive ductal carcinomas by a deep learning algorithm: the promise of AI support in radiology	Ueda D, Yamamoto A, Takashima T, Onoda N, Noda S, Kashiwagi S, Morisaki T, Tsutsumi S, Honjo T, Shimazaki A, Goto T and Miki Y	https://pubmed.ncbi.nlm.nih.gov/33200356/	2021	IDC	Japan	529	Not reported	Not reported	No
Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study	Kim HE, Kim HH, Han BK, Kim KH, Han K, Nam H, Lee EH and Kim EK	https://pubmed.ncbi.nlm.nih.gov/33334578/	2020		South Korea	Not reported	145,663	Both	No
An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization	Shen Y, Wu N, Phang J, Park J, Liu K, Tyagi S, Heacock L, Kim SG, Moy L, Cho K, Geras KJ	https://pubmed.ncbi.nlm.nih.gov/33383334/	2020	NYU breast cancer screening dataset	USA	141,472	1,001,093	Screening	No
Looking for Abnormalities in Mammograms With Self- and Weakly Supervised Reconstruction	Tardy M and Mateus D	https://pubmed.ncbi.nlm.nih.gov/33417539/	2021		Not reported	Not reported	2500	Not reported	No
Feature-Sensitive Deep Convolutional Neural Network for Multi-Instance Breast Cancer Detection	Wang Y, Zhang L, Shu X, Feng Y, Zhang Y and Lv Q	https://pubmed.ncbi.nlm.nih.gov/33600319/	2022		China	4627	10,518	Screening	Not reported
Multiview multimodal network for breast cancer diagnosis in	Song J, Zheng Y, Zakir Ullah M,	https://pubmed.ncbi.nlm.	2021		Not reported	95	760	Not reported	Not reported

(*continued on next page*)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
contrast-enhanced spectral mammography images	Wang J, Jiang Y, Xu C, Zou Z and Ding G	nih.gov/33966155/							
Application of deep learning in the detection of breast lesions with four different breast densities	Li H, Ye J, Liu H, Wang Y, Shi B, Chen J, Kong A, Xu Q and Cai J	https://pubmed.ncbi.nlm.nih.gov/34132495/	2021		China	608	Not reported	Not reported	Not reported
MommiNet-v2: Mammographic multi-view mass identification networks	Yang Z, Cao Z, Zhang Y, Tang Y, Lin X, Ouyang R, Wu M, Han M, Xiao J, Huang L, Wu S, Chang P and Ma J	https://pubmed.ncbi.nlm.nih.gov/34399154/	2021		China	2749	Not reported	Not reported	No
Anomaly Detection of Calcifications in Mammography Based on 11,000 Negative Cases	Hou R, Peng Y, Grimm LJ, Ren Y, Mazurowski MA, Marks JR, King LM, Maley CC, Hwang ES and Lo JY	https://pubmed.ncbi.nlm.nih.gov/34788216/	2022		USA	12,010	53,239	Not reported	No
Stand-Alone Use of Artificial Intelligence for Digital Mammography and Digital Breast Tomosynthesis Screening: A Retrospective Evaluation	Romero-Martín S, Elías-Cabot E, Raya-Povedano JL, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M	https://pubmed.ncbi.nlm.nih.gov/34904872/	2022		Spain	16,067	Not reported	Screening	No
Mammography Image-Based Diagnosis of Breast Cancer Using Machine Learning: A Pilot Study	Alshammari MM, Almuhanna A, Alhiyafi J	https://pubmed.ncbi.nlm.nih.gov/35009746/	2021		Saudi Arabia	42	Not reported	Not reported	No
Detection and Weak Segmentation of Masses in Gray-Scale Breast Mammogram Images Using Deep Learning	Kim YJ and Kim KG	https://pubmed.ncbi.nlm.nih.gov/35040607/	2022		South Korea	Not reported	300	Not reported	Not reported
Architectural Distortion-Based Digital Mammograms Classification Using Depth Wise Convolutional Neural Network	Rehman KU, Li J, Pei Y, Yasin A, Ali S, Saeed Y	https://pubmed.ncbi.nlm.nih.gov/35053013/	2021	PINUM	Pakistan	Not reported	289	Not reported	No
Diagnostic Performance of AI for Cancers Registered in A Mammography Screening Program: A Retrospective Analysis	Kizildag Yirgin I, Koyluoglu YO, Seker ME, Ozkan Gurdal S, Ozaydin AN, Ozcinar B, CabioÄŸlu N, Ozmen V and Aribal E	https://pubmed.ncbi.nlm.nih.gov/35060413/	2022		Turkey	Not reported	311	Screening	No
A High-Performance Deep Neural Network Model for BI-RADS Classification of Screening Mammography	Tsai KJ, Chou MC, Li HM, Liu ST, Hsu JH, Yeh WC, Hung CM, Yeh CY and Hwang SH	https://pubmed.ncbi.nlm.nih.gov/35161903/	2022		Taiwan	1490	5733	Screening	No
An Artificial Intelligence–based Mammography Screening Protocol for Breast Cancer: Outcome and Radiologist Workload	Lauritzen AD, RodrÃguez-Ruiz A, von Euler-Chelpin MC, Lynge E, Vejborg I, Nielsen M, Karssemeijer N and Lillholm M	https://pubmed.ncbi.nlm.nih.gov/35438561/	2022		Denmark	168,372	Not reported	Screening	No

(*continued on next page*)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques	Asma Baccouche, Begonya Garcia-Zapirain, Yufeng Zheng, Adel S Elmaghraby	https://pubmed.ncbi.nlm.nih.gov/35594582/	2022	University of Conneticut Center DigiMammo Database (UCHCDM)	USA	230	826	Screening	Yes
Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network	Zhang W, Doi K, Giger ML, Wu Y, Nishikawa RM and Schmidt RA	https://pubmed.ncbi.nlm.nih.gov/8058017/	1994		Not reported	Not reported	Not reported	Not reported	Yes
An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms	Zhang W, Doi K, Giger ML, Nishikawa RM and Schmidt RA	https://pubmed.ncbi.nlm.nih.gov/8860907/	1996		Not stated	Not reported	89	Not reported	Yes
Adequacy Testing of Training Set Sample Sizes in the Development of a Computer Assisted Diagnosis Scheme	Zheng B, Chang YH, Good WF and Gur D	https://pubmed.ncbi.nlm.nih.gov/9232169/	1996		USA	Not reported	618	Not reported	Yes
*Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms	Nagel RH, Nishikawa RM, Papaioannou J and Doi K	https://pubmed.ncbi.nlm.nih.gov/9725141/	1998	39 mammogram database 50 mammogram database	Not reported Not reported	Not reported Not reported	39 50	Not reported Not reported	Yes Yes
Automatic Breast Mass Segmentation and Classification Using Subtraction of Temporally Sequential Digital Mammograms	Kosmia Loizidou, Galateia Skouroumouni, Christos Nikolaou, Costas Pitris	https://pubmed.ncbi.nlm.nih.gov/36519002/	2022			62	320	Screening	No
Impact of a Categorical AI System for Digital Breast Tomosynthesis on Breast Cancer Interpretation by Both General Radiologists and Breast Imaging Specialists	Jiye G Kim, Bryan Haslam, Abdul Rahman Diab, Ashwin Sakhare, Giorgia Grisot, Hyunkwang Lee, Jacqueline Holt, Christoph I Lee, William Lotter, A Gregory Sorensen	https://pubmed.ncbi.nlm.nih.gov/38323914/	2024		USA	240	240	Screening	No
Interpretable Radiomic Signature for Breast Microcalcification Detection and Classification	Francesco Prinzi, Alessia Orlando, Salvatore Gaglio, Salvatore Vitabile	https://pubmed.ncbi.nlm.nih.gov/38351223/	2024		Italy	Not reported	161	Not reported	No
Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis.	Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L.	https://pubmed.ncbi.nlm.nih.gov/35750400/	2022		German	453,104	1,193,197	Screening	No
An integrated framework for breast mass classification and diagnosis using stacked ensemble of residual neural networks	Baccouche A, Garcia-Zapirain B, Elmaghraby AS. An integrated framework for breast mass classification and	https://pubmed.ncbi.nlm.nih.gov/35851592/	2022		Mexico	Not Reported	389	Not reported	Not reported

(*continued on next page*)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
	diagnosis using stacked ensemble of residual neural networks.								
Improving lesion detection in mammograms by leveraging a Cycle-GAN-based lesion remover	Lee J, Nishikawa RM	https://pubmed.ncbi.nlm.nih.gov/38303004/	2024		USA	4832	10,310	Screening	No
External Evaluation of a Mammography-based Deep Learning Model for Predicting Breast Cancer in an Ethnically Diverse Population.	Omoleye OJ, Woodard AE, Howard FM, Zhao F, Yoshimatsu TF, Zheng Y, Pearson AT, Levental M, Aribisala BS, Kulkarni K, Karczmar GS, Olopade OI, Abe H, Huo D.	https://pubmed.ncbi.nlm.nih.gov/38074785/	2023		USA	2096	6435	Screening	Not reported
*Breast density prediction from low and standard dose mammograms using deep learning: effect of image resolution and model training approach on prediction quality.	Squires S, Harkness EF, Mackenzie A, Evans DG, Howell SJ, Astley SM	https://pubmed.ncbi.nlm.nih.gov/38701765/	2024	ALDRAM PROCAS	Not reported UK	148 4051	888 15,290	Not reported Screening	Not reported Not reported
Identifying factors that indicate the possibility of non-visible cases on mammograms using mammary gland content ratio estimated by artificial intelligence	Kai C, Otsuka T, Nara M, Kondo S, Futamura H, Kodama N, Kasai S	https://pubmed.ncbi.nlm.nih.gov/38505584/	2024		Japan	Not Reported	211,897	Not reported	No
Development and validation of a deep learning model for detection of breast cancers in mammography from multi-institutional datasets	Ueda D, Yamamoto A, Onoda N, Takashima T, Noda S, Kashiwagi S, Morisaki T, Fukumoto S, Shiba M, Morimura M, Shimono T, Kageyama K, Tatekawa H, Murai K, Honjo T, Shimazaki A, Kabata D, Miki Y.	https://pubmed.ncbi.nlm.nih.gov/35324962/	2022		Japan	1901	6491	Screening	Yes
*AI for interpreting screening mammograms: implications for missed cancer in double reading practices and challenging-to-locate lesions	Jiang Z, Gandomkar Z, Trieu PDY, Taba ST, Barron ML, Lewis SJ.	https://pubmed.ncbi.nlm.nih.gov/38789575/	2024	Dataset 1 Dataset 2	Australia Australia	729 Not reported	Not reported Not reported	Screening Screening	Not reported Not reported
Effects of vitamin D supplementation on a deep learning-based mammographic evaluation in SWOG S0812	McGuinness JE, Anderson GL, Mutasa S, Hershman DL, Terry MB, Tehranifar P, Lew DL, Yee M, Brown EA, Kairouz SS, Kuwajerwala N, Bevers TB, Doster JE, Zarwan C, Kruper L, Minasian LM, Ford L, Arun B, Neuhouser ML,	https://pubmed.ncbi.nlm.nih.gov/38814817/	2024	SWOG S0812	USA	208	273	Clinical Trial	No

(*continued on next page*)

(continued)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
	Goodman GE, Brown PH, Ha R, Crew KD.								
Multi-modal artificial intelligence for the combination of automated 3D breast ultrasound and mammograms in a population of women with predominantly dense breasts	Tan T, Rodriguez-Ruiz A, Zhang T, Xu L, Beets-Tan RGH, Shen Y, Karssemeijer N, Xu J, Mann RM, Bao L.	https://pubmed.ncbi.nlm.nih.gov/36645507/	2023		China	430	Not reported	Symptomatic or voluntary self-motivated breast screening	No
Proposal and Definition of an Intelligent Clinical Decision Support System Applied to the Screening and Early Diagnosis of Breast Cancer	Casal-Guisande M, Álvarez-Pazó A, Cerqueiro-Pequeño J, Bouza-Rodríguez JB, Peláez-Lourido G, Comesaña-Campos A	https://pubmed.ncbi.nlm.nih.gov/36980595/	2023		USA	130		Not reported	Not reported
Temporal Machine Learning Analysis of Prior Mammograms for Breast Cancer Risk Prediction. Cancers	Li H, Robinson K, Lan L, Baughan N, Chan CW, Embury M, Whitman GJ, El-Zein R, Bedrosian I, Giger ML.	https://pubmed.ncbi.nlm.nih.gov/37046802/	2023		USA	99	318	Screening and Clinical Trial	Not reported
Reducing the number of unnecessary biopsies for mammographic BI-RADS 4 lesions through a deep transfer learning method	Meng M, Li H, Zhang M, He G, Wang L, Shen D.	https://pubmed.ncbi.nlm.nih.gov/37312026/	2023		Not Reported	1980	4330	Not reported	No
Unsupervised anomaly detection with generative adversarial networks in mammography	Park S, Lee KH, Ko B, Kim N.	https://pubmed.ncbi.nlm.nih.gov/36805637/	2023	No name provided	Asia	22,848	105,948	Screening	No
External Validation of a Mammography-Derived AI-Based Risk Model in a U.S. Breast Cancer Screening Cohort of White and Black Women	Gastounioti A, Eriksson M, Cohen EA, Mankowski W, Pantalone L, Ehsan S, McCarthy AM, Kontos D, Hall P, Conant EF.	https://pubmed.ncbi.nlm.nih.gov/36230723/	2022		USA	5139	Not Reported	Screening	No
Evaluating Recalibrating AI Models for Breast Cancer Diagnosis in a New Context: Insights from Transfer Learning, Image Enhancement and High-Quality Training Data Integration.	Jiang Z, Gandomkar Z, Trieu PDY, Tavakoli Taba S, Barron ML, Obeidy P, Lewis SJ.	https://pubmed.ncbi.nlm.nih.gov/38254813/	2024	Australian mammographic database Lifepool	Australia	Not Reported	1712	Screening	No
Impact of Tomosynthesis Acquisition on 3D Segmentations of Breast Outline and Adipose/Dense Tissue with AI: A Simulation-Based Study. Tomography.	Barufaldi B, Gomes J, Rego TGD, Malheiros Y, Filho TMS, Borges LR, Acciavatti RJ, Surti S, Maidment ADA.	https://pubmed.ncbi.nlm.nih.gov/37489471/	2023		Not Reported	660	Not Reported	Not Reported	No
Deep Learning Analysis of Mammography for Breast Cancer Risk Prediction in Asian Women.	Kim H, Lim J, Kim HG, Lim Y, Seo BK, Bae MS.	https://pubmed.ncbi.nlm.nih.gov/37443642/	2023		Not Reported	600	1023	Screening	No
Artificial Intelligence Improves Detection	Yoen H, Chang JM.	https://pubmed.	2023		Not Reported	413	Not Reported	Screening	No

(continued on next page)

(continued)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
of Supplemental Screening Ultrasound-detected Breast Cancers in Mammography. J Breast Cancer.		ncbi.nlm.nih.gov/37704383/							
Deep Learning vs Traditional Breast Cancer Risk Models to Support Risk-Based Mammography Screening	Lehman CD, Mercaldo S, Lamb LR, King TA, Ellisen LW, Specht M, Tamimi RM.	https://pubmed.ncbi.nlm.nih.gov/35876790/	2022	Breast Imaging Data Repository Archive	USA	57617	119139	Screening	No
A Deep Learning Decision Support Tool to Improve Risk Stratification and Reduce Unnecessary Biopsies in BI-RADS 4 Mammograms	Ezeana CF, He T, Patel TA, Kaklamani V, Elmi M, Brigmon E, Otto PM, Kist KA, Speck H, Wang L, Ensor J, Shih YT, Kim B, Pan IW, Cohen AL, Kelley K, Spak D, Yang WT, Chang JC, Wong STC.	https://pubmed.ncbi.nlm.nih.gov/38074778/	2023		USA	4209	Not reported	Not reported	No
End-to-End Calcification Distribution Pattern Recognition for Mammograms: An Interpretable Approach with GNN	Yao MM, Du H, Hartman M, Chan WP, Feng M.	https://pubmed.ncbi.nlm.nih.gov/35741186/	2022		Taiwan	292	584	Referrals from breast clinics or screening centres	No
A deep learning approach for virtual contrast enhancement in Contrast Enhanced Spectral Mammography.	Rofena A, Guarrasi V, Sarli M, Piccolo CL, Sammarra M, Zobel BB, Soda P.	https://pubmed.ncbi.nlm.nih.gov/38810487/	2024	CESM@UCBM	Italy	105	1138	Screening	No
Development and validation of an infrared-artificial intelligence software for breast cancer detection.	Martín-Del-Campo-Mena E, Sánchez-Méndez PA, Ruvalcaba-Limon E, Lazcano-Ramírez FM, Hernández-Santiago A, Juárez-Aburto JA, Larios-Cruz KY, Hernández-Gómez LE, Merino-González JA, González-Mejía Y.	https://pubmed.ncbi.nlm.nih.gov/37206999/	2023	No name provided	Mexico	3812	2699	Screening	Not reported
Developing a Supplementary Diagnostic Tool for Breast Cancer Risk Estimation Using Ensemble Transfer Learning.	Hanis TM, Ruhaiyem NIR, Arifin WN, Haron J, Wan Abdul Rahman WF, Abdullah R, Musa KI.	https://pubmed.ncbi.nlm.nih.gov/37238264/	2023		Malaysia	Not reported	7452	Screening	No
Deep learning performance for detection and classification of microcalcifications on mammography.	Pesapane F, Trentin C, Ferrari F, Signorelli G, Tantrige P, Montesano M, Cicala C, Virgoli R, D'Acquisto S, Nicosia L, Origgi D, Cassano E.	https://pubmed.ncbi.nlm.nih.gov/37934382/	2023		Europe (single institution)	1000	1986	Screening	No
Long-Term Performance of an Image-Based Short-Term Risk Model for Breast Cancer.	Eriksson M, Czene K, Vachon C, Conant EF, Hall P.	https://pubmed.ncbi.nlm.nih.gov/36930854/	2023	KARMA	Sweden	8604	2028	Screening	No
Artificial intelligence-supported screen	Lång K, Josefsson V, Larsson AM,	https://pubmed.	2023		China		178	Screening	Not reported

(continued on next page)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study.	Larsson S, Högberg C, Sartor H, Hofvind S, Andersson I, Rosso A.	ncbi.nlm.nih.gov/37541274/							
Attention-Based Deep Learning System for Classification of Breast Lesions-Multimodal, Weakly Supervised Approach.	Bobowicz M, Rygusik M, Buler J, Buler R, Ferlin M, Kwasigroch A, Szurowska E, Grochowski M.	https://pubmed.ncbi.nlm.nih.gov/37345041/	2023	MUG	Poland	789	1968	Screening	No
Challenge PRECISION Consortium Steering Group. Application of deep learning on mammographies to discriminate between low and high-risk DCIS for patient participation in active surveillance trials.	Alaeikhanehshir S, Voets MM, van Duijnhoven FH, Lips EH, Groen EJ, van Oirsouw MCJ, Hwang SE, Lo JY, Wesseling J, Mann RM, Teuwen J; Grand	https://pubmed.ncbi.nlm.nih.gov/38576031/	2024		Switzerland	464	681	Screening	No
Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study.	Lång K, Josefsson V, Larsson AM, Larsson S, Högberg C, Sartor H, Hofvind S, Andersson I, Rosso A.	https://pubmed.ncbi.nlm.nih.gov/37541274/	2023	MASAI	Sweden	80,033	Not reported	Screening	No
Impact of GAN artifacts for simulating mammograms on identifying mammographically occult cancer	Lee J, Mustafaev T, Nishikawa RM.	https://pubmed.ncbi.nlm.nih.gov/37840849/	2023	CGAN and MO	Not reported	1699	Not reported	Screening	No
*Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model.	Yala A, Mikhael PG, Strand F, Lin G, Satuluru S, Kim T, Banerjee I, Gichoya J, Trivedi H, Lehman CD, Hughes K, Sheedy DJ, Matthis LM, Karunakaran B, Hegarty KE, Sabino S, Silva TB, Evangelista MC, Caron RF, Souza B, Mauad EC, Patalon T, Handelman-Gotlib S, Guindy M, Barzilay R.	https://pubmed.ncbi.nlm.nih.gov/34767469/	2022	MGH Novant Emory Maccabi-Assuta Karolinska CGMH Barretos	USA USA USA Israel Sweden Taiwan Brazil	7005 1887 16,495 6189 7353 13,536 5900	25,855 14,157 44,008 6189 19,328 13,536 5900	Screening Screening Screening Screening Screening Screening Screening	Not reported Not reported Not reported Not reported Not reported Not reported Not reported
AI-Based Cancer Detection Model for	Jailin C, Mohamed S, Iordache R, Milioni De Carvalho	https://pubmed.ncbi.nlm.	2022		USA, UK, India, Egypt	1673	7443	Screening	Not reported

(*continued on next page*)

(continued)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
Contrast-Enhanced Mammography.	P, Ahmed SY, Abdel Sattar EA, Moustafa AFI, Gomaa MM, Kamal RM, Vancamberg L.	nih.gov/37627859/							
Impact of loss functions on the performance of a deep neural network designed to restore low-dose digital mammography	Shan H, Vimieiro RB, Borges LR, Vieira MAC, Wang G.	https://pubmed.ncbi.nlm.nih.gov/37316093/	2023		Brazil	100	Not reported	Screening	No
Improved PAA algorithm for breast mass detection in mammograms.	Liu W, Zeng P, Jiang J, Chen J, Chen L, Hu C, Jian W, Diao X, Wang X.	https://pubmed.ncbi.nlm.nih.gov/38744058/	2024		China	553	991	Not Reported	Not reported
Protocol for evaluating the fitness for purpose of an artificial intelligence product for radiology reporting in the BreastScreen New South Wales breast cancer screening programme.	Warner-Smith M, Ren K, Mistry C, Walton R, Roder D, Bhola N, McGill S, O'Brien TA.	https://pubmed.ncbi.nlm.nih.gov/38806433/	2024		Australia	626,851	658,207	Screening	No
Deep learning-based breast region extraction of mammographic images combining pre-processing methods and semantic segmentation supported by Deeplab v3.	Zhou K and Li W and Zhao D	https://pubmed.ncbi.nlm.nih.gov/36766450/	2023		Vietnam	357	731	Not Reported	Not reported
Independent evaluation of a multi-view multi-task convolutional neural network breast cancer classification model using Finnish mammography screening data.	Isosalo A, Inkinen SI, Turunen T, Ipatti PS, Reponen J, Nieminen MT.	https://pubmed.ncbi.nlm.nih.gov/37230016/	2023		Finland	22,739	49,654	Screening	No
Adaptive Machine Learning Approach for Importance Evaluation of Multimodal Breast Cancer Radiomic Features	Del Corso G, Germanese D, Caudai C, Anastasi G, Belli P, Formica A, Nicolucci A, Palma S, Pascali MA, Pieroni S, Trombadori C, Colantonio S, Franchini M, Molinaro S.	https://pubmed.ncbi.nlm.nih.gov/38478187/	2024	P.I.N·K study	Italy	66	Not reported	Diagnostic	No
Radiomic and deep learning characterization of breast parenchyma on full field digital mammograms and specimen radiographs: a pilot study of a potential cancer field effect	Baughan N, Li H, Lan L, Embury M, Yim I, Whitman GJ, El-Zein R, Bedrosian I, Giger ML	https://pubmed.ncbi.nlm.nih.gov/37426053/	2023	No name provided	USA	103	Not reported	Not reported	No
Exploiting Patch Sizes and Resolutions for Multi-Scale Deep Learning in Mammogram Image Classification	Quintana GI, Li Z, Vancamberg L, Mougeot M, Desolneux A, Muller S	https://pubmed.ncbi.nlm.nih.gov/37237603/	2023	GEHC	Not stated	Not reported	1539	Not reported	Not reported

(continued on next page)

(continued)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
An automated machine learning tool for breast cancer diagnosis for healthcare professionals.	Shaikh TA, Ali R.	https://pubmed.ncbi.nlm.nih.gov/36325422/	2021	No name provided	Not stated	160	160	Not reported	Not reported
Prediction of Breast Cancer using Machine Learning Approaches	Rabiei R, Ayyoubzadeh SM, Sohrabei S, Esmaeili M, Atashi A.	https://pubmed.ncbi.nlm.nih.gov/35698545/	2022	Motamed cancer institute (ACECR)	Iran	1290	Not reported	Not reported	Not reported
Prediction of Breast Cancer using Machine Learning Approaches.	Rangarajan K, Gupta A, Dasgupta S, Marri U, Gupta AK, Hari S, Banerjee S, Arora C.	https://pubmed.ncbi.nlm.nih.gov/35803985/	2022	No name provided	India	Not reported	5633	Diagnostic	No
Saliency of breast lesions in breast cancer detection using artificial intelligence.	Pertuz S, Ortega D, Suarez É, Cancino W, Africano G, Rinta-Kiikka I, Arponen O, Paris S, Lozano A.	https://pubmed.ncbi.nlm.nih.gov/37996504/	2023	No name provided	Finland	382	1582	Screening	No
Evaluating the Margins of Breast Cancer Tumors by Using Digital Breast Tomosynthesis with Deep Learning: A Preliminary Assessment	Shia WC, Kuo YH, Hsu FR, Lin J, Wu WP, Wu HK, Yeh WC, Chen DR.	https://pubmed.ncbi.nlm.nih.gov/38786329/	2024	No name provided	Not stated, Europe, North America, Africa and Asia	2023	9073 image pairs	Not reported	No
Deep learning-based breast region segmentation in raw and processed digital mammograms: generalization across views and vendors.	Verboom SD, Caballo M, Peters J, Gommers J, van den Oever D, Broeders MJM, Teuwen J, Sechopoulos I.	https://pubmed.ncbi.nlm.nih.gov/38162417/	2024	No name provided	Not stated	808	3841	Not reported	No
BRAIxDet: Learning to detect malignant breast lesion with incomplete annotations	Chen Y, Liu Y, Wang C, Elliott M, Kwok CF, Peña-Solorzano C, Tian Y, Liu F, Frazer H, McCarthy DJ, Carneiro G.	https://pubmed.ncbi.nlm.nih.gov/38810516/	2024	Annotated Digital Mammograms and Associated Non-Image data (ADMANI)	Australia	771,542 exams	1,543,084	Screening	No
Fair evaluation of federated learning algorithms for automated breast density classification: The results of the 2022 ACR-NCI-NVIDIA federated learning challenge	Schmidt K, Bearce B, Chang K, Coombs L, Farahani K, Elbatel M, Mouheb K, Marti R, Zhang R, Zhang Y, Wang Y, Hu Y, Ying H, Xu Y, Testagrose C, Demirer M, Gupta V, Akünal Ü, Bujotzek M, Maier-Hein KH, Qin Y, Li X, Kalpathy-Cramer J, Roth HR	https://pubmed.ncbi.nlm.nih.gov/38776844/	2023	No name provided	China	2496	19,968	Screening	No
Automatic Calcification Morphology and Distribution Classification for Breast Mammograms With Multi-Task Graph Convolutional Neural Network	Du H, Yao MM, Liu S, Chen L, Chan WP, Feng M	https://pubmed.ncbi.nlm.nih.gov/37027577/	2023	Taipei Medical University (TMU) dataset	Taiwan	200	387	Not reported	No
Impact of Artificial Intelligence System and Volumetric Density on Risk Prediction of Interval, Screen-Detected, and	Vachon CM, Scott CG, Norman AD, Khanani SA, Jensen MR, Hruska CB, Brandt KR, Winham SJ, Kerlikowske K.	https://pubmed.ncbi.nlm.nih.gov/37104728/	2023	San Francisco Mammography Registry (SFMR) and Mayo clinic screening data	USA	7407 (2412 cases, 4995 controls)	Not reported	Screening	No

(continued on next page)

(continued)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
Advanced Breast Cancer.									
Investigating the detection of breast cancer with deep transfer learning using ResNet18 and ResNet34	Subaar C, Addai FT, Addison ECK, Christos O, Adom J, Owusu-Mensah M, Appiah-Agyei N, Abbey S.	https://pubmed.ncbi.nlm.nih.gov/38599202/	2024	National Radiological Society's Archive	Not stated	Not reported	1200	Not reported	Not reported
Automated classification of fat-infiltrated axillary lymph nodes on screening mammograms.	Song Q, diFlorio-Alexander RM, Sieberg RT, Dwan D, Boyce W, Stumetz K, Patel SD, Karagas MR, MacKenzie TA, Hassanpour S.	https://pubmed.ncbi.nlm.nih.gov/37751215/	2023	No name provided	Not stated	Not reported	886	Screening	No
Impact of AI for Digital Breast Tomosynthesis on Breast Cancer Detection and Interpretation Time.	Park EK, Kwak S, Lee W, Choi JS, Kooi T, Kim EK.	https://pubmed.ncbi.nlm.nih.gov/38568095/	2024	No name provided	USA	Not reported	2206	Diagnostic	No
A machine learning model based on readers' characteristics to predict their performances in reading screening mammograms.	Gandomkar Z, Lewis SJ, Li T, Ekpo EU, Brennan PC	https://pubmed.ncbi.nlm.nih.gov/35122217/	2022	No name provided	Not stated	535	Not reported	Screening	No
Implications for downstream workload based on calibrating an artificial intelligence detection algorithm by standalone-reader or combined-reader sensitivity matching	Dembrower K, Salim M, Eklund M, Lindholm P, Strand F.	https://pubmed.ncbi.nlm.nih.gov/37035276/	2023	No name provided	Sweden	6708	Not reported	Screening	Not reported
Monitoring Methodology for an AI Tool for Breast Cancer Screening Deployed in Clinical Centers	Aguilar C, Pacilè S, Weber N, Fillard P.	https://pubmed.ncbi.nlm.nih.gov/36836797/	2023	No name provided	USA	Not reported	36,581 records	Not reported	No
A Semiautonomous Deep Learning System to Reduce False Positives in Screening Mammography.	Pedemonte S, Tsue T, Mombourquette B, Truong Vu YN, Matthews T, Morales Hoil R, Shah M, Ghare N, Zingman-Daniels N, Holley S, Appleton CM, Su J, Wahl RL.	https://pubmed.ncbi.nlm.nih.gov/38597785/	2024	No name provided	USA, UK	54,769	163,828	Screening	No
Study on the differential diagnosis of benign and malignant breast lesions using a deep learning model based on multimodal images.	Du Y, Wang D, Liu M, Zhang X, Ren W, Sun J, Yin C, Yang S, Zhang L.	https://pubmed.ncbi.nlm.nih.gov/38687933/	2024	No name provided	Not stated	132	Not reported	Diagnostic	No
Artificial intelligence-based computer-assisted detection/diagnosis (AI-CAD) for screening mammography: Outcomes of AI-CAD in the mammographic interpretation workflow	Yoon JH, Han K, Suh HJ, Youk JH, Lee SE, Kim EK.	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10362167/	2023	No name provided	South Korea	5288	6499	Screening	No

(continued on next page)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
VAI-B: a multicenter platform for the external validation of artificial intelligence algorithms in breast imaging.	Cossío F, Schurz H, Engström M, Barck-Holst C, Tsirikoglou A, Lundström C, Gustafsson H, Smith K, Zackrisson S, Strand F.	https://pubmed.ncbi.nlm.nih.gov/36949901/	2023	No name provided	Sweden	17,859	35,575	Screening	No
Analysis of Specimen Mammography with Artificial Intelligence to Predict Margin Status.	Chen KA, Kirchoff KE, Butler LR, Holloway AD, Kapadia MR, Kuzmiak CM, Downs-Canner SM, Spanheimer PM, Gallagher KK, Gomez SM	https://pubmed.ncbi.nlm.nih.gov/37563337/	2023	No name provided	USA	Not reported	806	Diagnostic	No
A deep learning approach to estimate x-ray scatter in digital breast tomosynthesis: From phantom models to clinical applications.	Pinto MC, Mauter F, Michielsen K, Biniazan R, Kappler S, Sechopoulos I.	https://pubmed.ncbi.nlm.nih.gov/37394837/	2023	No name provided	Not stated	Not reported	600 digitally generated phantoms (not real), 5 DBT images from unknown number of patients	600 Digital Breast Phantoms used, only 5 clinical cases used	No
Diagnostic accuracy of automated ACR BI-RADS breast density classification using deep convolutional neural networks	Sexauer R, Hejduk P, Borkowski K, Ruppert C, Weikert T, Dellas S, Schmidt N.	https://pubmed.ncbi.nlm.nih.gov/36856841/	2023	No name provided	Switzerland	1665	4605	Both	No
An Artificial Intelligence-based Mammography Screening Protocol for Breast Cancer: Outcome and Radiologist Workload. Radiology.	Lauritzen AD, Rodríguez-Ruiz A, von Euler-Chelpin MC, Lynge E, Vejborg I, Nielsen M, Karssemeijer N, Lillholm M.	https://pubmed.ncbi.nlm.nih.gov/35438561/	2022	No name provided	Denmark	114,421	114,421	Screening	No
Deep learning analysis of contrast-enhanced spectral mammography to determine histoprognostic factors of malignant breast tumours.	Dominique C, Callonnec F, Berghian A, Defta D, Vera P, Modzelewski R, Decazes P.	https://pubmed.ncbi.nlm.nih.gov/35094119/	2022	No name provided	France	389	2460	Both	No
Using Deep Neural Network Approach for Multiple-Class Assessment of Digital Mammography.	Hsu SY, Wang CY, Kao YK, Liu KY, Lin MC, Yeh LR, Wang YM, Chen CI, Kao FC.	https://pubmed.ncbi.nlm.nih.gov/36553906/	2022	No name provided	Taiwan	150	5400	Screening	No
External Validation of an Ensemble Model for Automated Mammography Interpretation by Artificial Intelligence.	Hsu W, Hippe DS, Nakhaei N, Wang PC, Zhu B, Siu N, Ahsen ME, Lotter W, Sorensen AG, Naeim A, Buist DSM, Schaffter T, Guinney J, Elmore JG, Lee CI.	https://pubmed.ncbi.nlm.nih.gov/36409497/	2022	No name provided	USA	26,817	37,317	Screening	No
Transformer-based Deep Neural Network for Breast Cancer Classification on Digital Breast Tomosynthesis Images.	Lee W, Lee H, Lee H, Park EK, Nam H, Kooi T.	https://pubmed.ncbi.nlm.nih.gov/37293346/	2023	No name provided	USA	Not reported	6829	Screening	No
Virtual Biopsy by Using Artificial Intelligence-based	Barros V, Tlusty T, Barkan E, Hexter E,	https://pubmed.ncbi.nlm.	2022	No name provided	Israel and USA	9234	26,569	Diagnostic	No

(*continued on next page*)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
Multimodal Modeling of Binational Mammography Data.	Gruen D, Guindy M, Rosen-Zvi M	nih.gov/36283109/							
Multi-scale cascaded networks for synthesis of mammogram to decrease intensity distortion and increase model-based perceptual similarity.	Jiang G, He Z, Zhou Y, Wei J, Xu Y, Zeng H, Wu J, Qin G, Chen W, Lu Y	https://pubmed.ncbi.nlm.nih.gov/36196045/	2023	2023	Not stated	1646 cases	Not reported	Not reported	No
Automatic mammographic breast density classification in Chinese women: clinical validation of a deep learning model.	Lin X, Wu S, Li L, Ouyang R, Ma J, Yi C, Tang Y.	https://pubmed.ncbi.nlm.nih.gov/36683330/	2023	No name provided	China	3732 cases	42,152	Screening	Not reported
Effect of artificial intelligence-based computer-aided diagnosis on the screening outcomes of digital mammography: a matched cohort study.	Kim H, Choi JS, Kim K, Ko ES, Ko EY, Han BK.	https://pubmed.ncbi.nlm.nih.gov/37188881/	2023	No name provided	South Korea	3158	Not reported	Screening	No
A multi-stage fusion framework to classify breast lesions using deep learning and radiomics features computed from four-view mammograms.	Jones MA, Sadeghipour N, Chen X, Islam W, Zheng B.	https://pubmed.ncbi.nlm.nih.gov/37083190/	2023	No name provided	USA	1065 cases	Not reported	Screening	No
Quantitative evaluation of Saliency-Based Explainable artificial intelligence (XAI) methods in Deep Learning-Based mammogram analysis.	Cerekci E, Alis D, Denizoglu N, Camurdan O, Ege Seker M, Ozer C, Hansu MY, Tanyel T, Oksuz I, Karaarslan E.	https://pubmed.ncbi.nlm.nih.gov/38364587/	2024	No name provided	Not stated	1496	Not reported	Screening	No
Prediction of Disease-Free Survival in Breast Cancer using Deep Learning with Ultrasound and Mammography: A Multicenter Study.	Han J, Hua H, Fei J, Liu J, Guo Y, Ma W, Chen J.	https://pubmed.ncbi.nlm.nih.gov/38281863/	2024	No name provided	Not stated	1242	Not reported	Screening	Not reported
Visual and quantitative evaluation of microcalcifications in mammograms with deep learning-based super-resolution	Honjo T, Ueda D, Katayama Y, Shimazaki A, Jogo A, Kageyama K, Murai K, Tatekawa H, Fukumoto S, Yamamoto A, Miki Y.	https://pubmed.ncbi.nlm.nih.gov/35834858/	2022	MedCity21	Japan	93	Not reported	Screening	No
Reliable quality assurance of X-ray mammography scanner by evaluation the standard mammography phantom image using an interpretable deep learning model.	Oh JH, Kim HG, Lee KM, Ryu CW.	https://pubmed.ncbi.nlm.nih.gov/35691109/	2022	Korean Institute for Accreditation of Medical Imaging (KIAMI)	South Korea	Not reported	2208	Not reported	No

(*continued on next page*)

(*continued*)

Article title	List of authors	Article DoI	Year of publication	Dataset name	Country of origin	Number of individuals	Number of images	Screening or diagnostic	Digitised from film images?
Artificial Intelligence for Breast Cancer Detection on Mammography: Factors Related to Cancer Detection.	Yoen H, Jang MJ, Yi A, Moon WK, Chang JM	https://pubmed.ncbi.nlm.nih.gov/38216413/	2024	No name provided	South Korea	1001	Not reported	Screening	Not reported
Patient Characteristics Impact Performance of AI Algorithm in Interpreting Negative Screening Digital Breast Tomosynthesis Studies.	Nguyen DL, Ren Y, Jones TM, Thomas SM, Lo JY, Grimm LJ.	https://pubmed.ncbi.nlm.nih.gov/38771177/	2024	No name provided	USA	4855	Not reported	Screening	No
Fair evaluation of federated learning algorithms for automated breast density classification: The results of the 2022 ACR-NCI-NVIDIA federated learning challenge	Schmidt K, Bearce B, Chang K, Coombs L, Farahani K, Elbatel M, Mouheb K, Marti R, Zhang R, Zhang Y, Wang Y, Hu Y, Ying H, Xu Y, Testagrose C, Demirer M, Gupta V, Akünal Ü, Bujotzek M, Maier-Hein KH, Qin Y, Li X, Kalpathy-Cramer J, Roth HR.	https://pubmed.ncbi.nlm.nih.gov/38776844/	2024	Digital Mammographic Imaging Screening Trial (DMIST)	Not stated	Not reported	103,890	Screening	No
Ipsilateral Lesion Detection Refinement for Tomosynthesis.	Ren Y, Liu X, Ge J, Liang Z, Xu X, Grimm LJ, Go J, Marks JR, Lo JY.	https://pubmed.ncbi.nlm.nih.gov/37227903/	2023	DBT dataset	USA, Europe	8034 cases	Not reported	Screening	No
Context attention pyramid network for computer-aided detection of microcalcification clusters in digital breast tomosynthesis.	Wang J, Sun H, Jiang K, Cao W, Chen S, Zhu J, Yang X, Zheng J.	https://pubmed.ncbi.nlm.nih.gov/37783114/	2023	No name provided	China	324	648	Screening	No
Impact of Transfer Learning Using Local Data on Performance of a Deep Learning Model for Screening Mammography	Condon JJJ, Trinh V, Hall KA, Reintals M, Holmes AS, Oakden-Rayner L, Palmer LJ.	https://pubmed.ncbi.nlm.nih.gov/38717291/	2024	BreastScreen SA (BSSA)	Australia	9185	Not reported	Screening	No

Appendix 4. Race and ethnicity categories reported within datasets

Datasets reporting Race	Category	%
DDSM (n = 2620)		Cohort 1 (Cohort 2)
	Asian	2.06 (0.2)
	Black	4.12 (20.4)
	Spanish Surname	6.55 (1.8)
	American Indian	0.00 (0.1)
	Other	0.75 (0.1)
	Unknown	30.34 (0.3)
	White	56.18 (77.0)
BCS-DBT (n = 5060)	American Indian or Alaskan Native	0.2
	Asian	3.6
	Black or African American	18.9
	Native Hawaiian or other Pacific Islander	<0.1
	White	73.1
	Other	1.0
	≥2 Races	1.1
	Not reported, declined, or unavailable	2.0
A Deep Learning Model to Triage Screening Mammograms: A Simulation Study (n = 80,818)	African American	Not reported
	Asian or Pacific Islander	

(*continued on next page*)

(*continued*)

Datasets reporting Race	Category	%
	White	
	Other	
A similarity measure method fusing deep feature for mammogram retrieval (number of individuals not reported)	Northeastern Chinese	100
Breast microcalcifications detection based on fusing features with DTCWT (n = 100)	Northeastern Chinese	100
Maghsoudi et al. ds1 (n = 2200)	African American/Black	45
	Caucasian/White	45
	Other	10
Maghsoudi et al. ds3a (n = 1662)	Caucasian/White	98
	Other	2
Maghsoudi et al. ds3b (n = 575)	African American/Black	53
	Caucasian/White	47
Maghsoudi et al. ds5 (n = 1592)	Caucasian/White	97
	Other	3
External Validation of a Mammography-Derived AI-Based Risk Model in a U.S. Breast Cancer Screening Cohort of White and Black Women (n = 5139)	Black	51
	White	42
	Other	6.7
	Missing	<1
Impact of Tomosynthesis Acquisition on 3D Segmentations of Breast Outline and Adipose/Dense Tissue with AI: A Simulation-Based Study (n = 660)	Asian	4
	Black	47
	White	48
	Other	1
MGH	African American	5
	Asian/Pacific Islander	5
	Hispanic	1
	White	81
	Other	6
Novant	African American	19
	American Indian or Alaskan Native	<1
	Asian	2
	Hispanic	3
	White	75
EMBED_Open_Data	African American	45
	American Indian or Alaskan Native	<1
	Asian	5
	Native Hawaiian or Other Pacific	1
	Islander:	45
	White	<1
	Multiple	
San Francisco Mammography Registry (SFMR) and Mayo clinic screening data (n = 2412)	Asian	16
	Black	3
	Hispanic	2
	Native American	0
	Other/Mixed	4
	Unknown	0
	White	76
Transformers Improve Breast Cancer Diagnosis from Unregistered Multi-View Mammograms (n = 949)	African American	Not reported
	Asian	
	Caucasian	
	Hispanic	
	Other	
Digital Mammographic Imaging Screening Trial (DMIST) (n = 4904)	American Indian or Alaska	<1
	Asian	2
	Black or African American	14
	Hispanic or Latino	4
	Other	1
	Unknown	<1
	White	80

Datasets reporting Ethnicity	Category	%
OMI-DB (*n* > 170,000)	Asian or Asian British - Indian Asian or Asian British - Pakistani Asian or Asian British - Bangladeshi Asian or Asian British - Any other Asian background Black or Black British - Caribbean Black or Black British - African Black or Black British - Any other Black background Mixed – White and Black Caribbean Mixed - White and Black African Mixed - White and Asian Mixed - Any other mixed background White - British	Not available prior to data transfer, around 20 % of participating data contributors do not collect this data

(*continued on next page*)

(*continued*)

Datasets reporting Ethnicity	Category	%
	White - Irish	
	White - any other background	
	Other Ethnic Groups - Chinese	
	Other Ethnic Groups - Any other ethnic group	
	Not stated	
PROCAS (n = 4051)	Asian/British Asian	2
	Black/Black British	1
	White	91
	Unknown	4

Datasets reporting Race and Ethnicity or Race/Ethnicity	Category	%
EMBED (22382)	*Race*	
	African American	40
	Asian	7
	White	40
	Ethnicity	
	Hispanic	6
Impact of a Categorical AI System for Digital Breast Tomosynthesis on Breast Cancer Interpretation by Both General Radiologists and Breast Imaging Specialists (n = 240)	*Race/Ethnicity*	7
	Asian	137
	Black or African	11
	American	75
	Hispanic	10
	White	
	Other	
External Evaluation of a Mammography-based Deep Learning Model for Predicting Breast Cancer in an Ethnically Diverse Population (n = 2096)	*Self-reported Race and*	
	Ethnicity	43
	African American	<1
	Alaska Native	4
	Asian or Pacific Islander	3
	Hispanic	41
	White	6
	Unknown or Missing data	
SWOG S0812 (n = 208)	*Race*	
	Black	6
	Native American	<1
	Multiracial	2
	White	77
	Unknown	1
	Ethnicity	
	Hispanic: 8 %	8
A Deep Learning Decision Support Tool to Improve Risk Stratification and Reduce Unnecessary Biopsies in BI-RADS 4 Mammograms (n = 4209)	*Race/Ethnicity*	
	African American	14
	Asian	7
	Caucasian (Non-	60
	Hispanic)	13
	Hispanic	<1
	Native American	6
	Other	1
	Unknown	
External Validation of an Ensemble Model for Automated Mammography Interpretation by Artificial Intelligence (n = 26,817)	*Race/Ethnicity*	
	Asian 3338	13
	Black 2972	11
	Hispanic 3699	14
	White 20,602	77
	Other 4093	15
Patient Characteristics Impact Performance of AI Algorithm in Interpreting Negative Screening Digital Breast Tomosynthesis Studies (n = 4855)	*Race*	
	Asian: 1351	26
	Black: 1261	26
	White: 1316	27
	Ethnicity	19
	Hispanic: 927	
Protocol for evaluating the fitness for purpose of an artificial intelligence product for radiology reporting in the BreastScreen New South Wales breast cancer screening programme (n = 626,851)	Not reported	Not reported
A Semiautonomous Deep Learning System to Reduce False Positives in Screening Mammography (n = 54,769)	Asian	Not
	Black	reported
	Hispanic	
	White	
	Other	

(*continued on next page*)

(*continued*)

Datasets reporting Race and Ethnicity or Race/Ethnicity	Category	%
Analysis of Specimen Mammography with Artificial Intelligence to Predict Margin Status	Asian Hispanic Non-Hispanic Black Non-Hispanic white Other/Unknown	Not reported

Appendix 5. Completeness of metadata reporting for the accessible data

Y: Yes, the metadata category has been reported within the dataset documentation
N: No, the metadata category has not been reported within the dataset documentation

Dataset name	Country of origin	Publication date	Source of data (screening/diagnostic)	Labels	Image format	Number of individuals	Number of images	Ethical approval	Consent/Waived consent	Version number	Licensing arrangements	Funding	Includes advice from advisory boards or patient participation groups	Highlights known missing groups	Proportion of attributes recorded as 'unknown' or other	Age	Sex/gender	Race	Ethnicity	Socioeconomic status	Data modifications
EMBED_Open_Data (Emory Breast Imaging Dataset)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y
Breast Cancer Screening - Digital Breast Tomosynthesis (Breast-Cancer-Screening-DBT)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	N
InBreast	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	N	Y	N	N	N	N	N
Breast Micro-Calcifications Dataset with Precisely AnNtated Sequential Mammograms	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	N	N	N	Y
King Abdulaziz University Mammogram Dataset (KAU-BCMD)	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	N	N	N	N	Y
OPTIMAM	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	N	Y	Y	N	Y	N	N
Categorized Digital Database for Low energy and Subtracted Contrast Enhanced Spectral Mammography images (CDD-CESM)	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	N	Y	Y	N	N	N	N
VinDr	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	N	N	Y	N	N	N	N	N
Breast Cancer Digital Repository (Iberian)	Y	Y	N	Y	Y	Y	Y	N	Y	Y	N	N	N	N	N	Y	Y	N	N	N	N
Mammographic Mass	Y	Y	Y	N	Y	N	Y	N	N	N	N	N	Y	Y	N	Y	N	N	N	N	Y
DDSM: Digital Database for Screening Mammography	Y	Y	Y	Y	Y	Y	N	N	N	N	N	Y	N	N	N	Y	Y	Y	N	N	N
Transformers Improve Breast Cancer Diagnosis from Unregistered	N	Y	N	Y	N	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y	Y	N	N	N

Multi-View Mammograms (Chen et al)																					
Ambra UNIFESP Mammography dataset	Y	Y	N	Y	Y	Y	Y	Y	N	N	Y	N	N	N	N	Y	Y	N	N	N	N
Scottish Breast Screening Service (SBSS)	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	N	N	N	Y	Y	N	N	Y	N
Breast Micro-Calcifications Dataset with Precisely Annotated Sequential Mammograms	Y	Y	N	Y	Y	Y	Y	Y	N	Y	N	Y	N	N	N	Y	Y	N	N	N	N
Digital mammography Dataset for Breast Cancer Diagnosis Research (DMID)	Y	Y	Y	Y	Y	N	Y	Y	N	Y	N	N	N	N	N	N	N	N	N	N	N
VICTRE	N	Y	Y	Y	Y	Y	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
Mammogram Mastery: A Robust Dataset for Breast Cancer Detection and Medical Education	Y	Y	N	Y	Y	N	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N	N
MIAS Mammography	Y	Y	N	N	Y	N	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	Y
The Chinese Mammography Database (CMMD)	Y	Y	N	N	Y	Y	Y	N	N	N	N	N	N	N	N	Y	N	N	N	N	N
Mammograms-Breast Cancer Images	N	Y	N	N	Y	N	Y	N	N	Y	N	N	N	N	N	Y	Y	N	N	N	N
CSAW-S	N	Y	N	N	Y	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N
Mammogram Density Assessment Dataset	N	Y	N	Y	Y	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N
Breast Mammography Image Dataset with Masses	N	Y	N	N	Y	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Seoul National University Bundang Hospital Mammographic Database (SNUBH-MDB-mCi)	N	Y	N	N	Y	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
BIRAD Mammo	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N
RSNA Mammography Breast Cancer Detection PNG	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Full mammogram	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

. (*continued*).

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

References

1 Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209–49. https://doi.org/10.3322/CAAC.21660.

2 Morton R, Sayma M, Sura MS. Economic analysis of the breast cancer screening program used by the UK NHS: Should the program be maintained? Breast Cancer 2017;9:217–25. https://doi.org/10.2147/BCTT.S123558.

3. Royal College of Radiologists. Clinical Radiology Workforce Census. 2022.

4 Barrios CH. Global challenges in breast cancer detection and treatment. Breast 2022;62:S3–6. https://doi.org/10.1016/J.BREAST.2022.02.003.

5 Lång K, Josefsson V, Larsson AM, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol 2023;24(8):936–44. https://doi.org/10.1016/S1470-2045(23)00298-X.

6 Yoon JH, Strand F, Baltzer PAT, et al. Standalone AI for breast cancer detection at screening digital mammography and digital breast tomosynthesis: a systematic review and meta-analysis. Radiology 2023;307(5). https://doi.org/10.1148/RADIOL.222639/ASSET/IMAGES/LARGE/RADIOL.222639.VA.JPEG.

7 Jairam MP, Ha R. A review of artificial intelligence in mammography. Clin Imaging 2022;88:36–44. https://doi.org/10.1016/J.CLINIMAG.2022.05.005.

8. Data Science Institute. AI Central. American College of Radiology. Accessed July 18, 2023, https://aicentral.acrdsi.org/; 2023.
9 Nalla V, Pouriyeh S, Parizi RM, et al. Deep learning for computer-aided abnormalities classification in digital mammogram: a data-centric perspective. Curr Probl Diagn Radiol 2024;53(3):346–52. https://doi.org/10.1067/J.CPRADIOL.2024.01.007.
10 Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health 2021;3(4): e260–5. https://doi.org/10.1016/S2589-7500(20)30317-4.
11 Logan J, Kennedy PJ, Catchpoole D. A review of the machine learning datasets in mammography, their adherence to the FAIR principles and the outlook for the future. Sci Data 2023;10(1):1–11. https://doi.org/10.1038/s41597-023-02430-6.
12 Wen D, Khan SM, Xu AJ, et al. Characteristics of publicly available skin cancer image datasets: a systematic review. Lancet Digit Health 2022;4(1):e64–74. https://doi.org/10.1016/S2589-7500(21)00252-1.
13 Khan SM, Liu X, Nath S, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. Lancet Digit Health 2021;3(1):e51–66. https://doi.org/10.1016/S2589-7500(20)30240-5.
14 Alderman JE, Charalambides M, Sachdeva G, et al. Revealing transparency gaps in publicly available COVID-19 datasets used for medical artificial intelligence development—a systematic review. Lancet Digit Health 2024;6(11):e827–47. https://doi.org/10.1016/S2589-7500(24)00146-8.
15 Daneshjou R, Vodrahalli K, Novoa RA, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv 2022;8(32). https://doi.org/10.1126/SCIADV.ABQ6147.
16 Clarke CN, Cortina CS, Fayanju OM, et al. Breast Cancer Risk and Screening in Transgender Persons: A Call for Inclusive Care. Published online. 2021. https://doi.org/10.1245/s10434-021-10217-5.
17 Hirko KA, Rocque G, Reasor E, et al. The impact of race and ethnicity in breast cancer—disparities and implications for precision oncology. BMC Med 2022;20(1): 1–12. https://doi.org/10.1186/S12916-022-02260-0/TABLES/2.
18 Ganapathi S, Palmer J, Alderman JE, et al. Tackling bias in AI health datasets through the STANDING Together initiative. Nat Med 2022;28(11):2232–3. https://doi.org/10.1038/s41591-022-01987-w.
19 Arora A, Alderman JE, Palmer J, et al. The value of standards for health datasets in artificial intelligence-based applications. Nat Med 2023;10:36. https://doi.org/10.1038/s41591-023-02608-w.
20 Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1). https://doi.org/10.1186/S13643-016-0384-4.
21. STANDING Together Working Group. Draft Recommendations for Healthcare Dataset Standards Supporting Diversity, Inclusivity and Generalisability. https://www.datadiversity.org/draft-standards. [Accessed 20 June 2023].
22. NHS Data Model and Dictionary. 16 + 1 ethnic data categories. https://www.datadictionary.nhs.uk/data_elements/ethnic_category.html. [Accessed 20 November 2023].
23 Ramadan SZ. Using convolutional neural network with cheat sheet and data augmentation to detect breast cancer in mammograms. Comput Math Methods Med 2020:2020. https://doi.org/10.1155/2020/9523404.
24 Altaf MM. A hybrid deep learning model for breast cancer diagnosis based on transfer learning and pulse-coupled neural networks. Math Biosci Eng 2021;18(5):5029–46. https://doi.org/10.3934/MBE.2021256.
25 Zahoor S, Shoaib U, Lali IU. Breast cancer mammograms classification using deep neural network and entropy-controlled whale optimization algorithm. Diagnostics (Basel) 2022;12(2). https://doi.org/10.3390/DIAGNOSTICS12020557.
26 San Miguel Y, Gomez SL, Murphy JD, et al. Age-related differences in breast cancer mortality according to race/ethnicity, insurance, and socioeconomic status. BMC Cancer 2020;20(1):1–9. https://doi.org/10.1186/S12885-020-6696-8/TABLES/3.
27 Lett E, Asabor E, Beltrán S, Cannon AM, Arah OA. Conceptualizing, contextualizing, and operationalizing race in quantitative health sciences research. Ann Fam Med 2022;20(2):157–63. https://doi.org/10.1370/AFM.2792/-/DC1.
28 Cerdeña JP, Plaisime MV, Tsai J. From race-based to race-conscious medicine: how anti-racist uprisings call us to act. Lancet 2020;396(10257):1125–8. https://doi.org/10.1016/S0140-6736(20)32076-6.
29. American Medical Association, Center for Health Justice. Advancing Health Equity: A Guide to Language, Narrative and Concepts. 2021.
30 Gabbert W. Concepts of Ethnicity. Lat Am Caribb Ethn Stud 2006;1(1):85–103. https://doi.org/10.1080/17486830500510034.
31 Monteiro Cordeiro N, Facina G, Pinto Nazário AC, et al. Towards precision medicine in breast imaging: a novel open mammography database with tailor-made 3D image retrieval for AI and teaching. Comput Methods Programs Biomed 2024:248. https://doi.org/10.1016/J.CMPB.2024.108117.
32 Scheim AI, Baker KE, Restar AJ, Sell RL. Health and Health Care Among Transgender Adults in the United States. https://doi.org/101146/annurev-publhealth-052620-10031343; 2022. p. 503–23. https://doi.org/10.1146/ANNUREV-PUBLHEALTH-052620-100313.
33 Nguyen DL, Ren Y, Jones TM, Thomas SM, Lo JY, Grimm LJ. Patient characteristics impact performance of AI algorithm in interpreting negative screening digital breast tomosynthesis studies. Radiology 2024;311(2). https://doi.org/10.1148/RADIOL.232286/ASSET/IMAGES/LARGE/RADIOL.232286.VA.JPEG.
34 Southam CM, Lacks H, Chester SK, Southam M, Kettering S, Vernon LF. Tuskegee syphilis study not America's only medical scandal: Chester M. Southam, MD, Henrietta Lacks, and the Sloan-Kettering research scandal. J Health Ethics 2020;16 (2):3. https://doi.org/10.18785/ojhe.1602.03.
35 Fantz CR. Barriers to quality health care for the transgender population. Clin Biochem 2014;47(10−11):983–7. https://doi.org/10.1016/J.CLINBIOCHEM.2014.02.009.
36 Macpherson MS, Hutchinson CE, Horst C, Goh V, Montana G. Patient reidentification from chest radiographs: aninterpretable deep metric learning approach and its applications. Radiol Artif Intell 2023;5(6). https://doi.org/10.1148/RYAI.230019/ASSET/IMAGES/LARGE/RYAI.230019.VA.JPEG.
37 Sonnenblick EB, Lebron-Zapata L, Yang R, et al. Breast imaging for transgender individuals: assessment of current practice and needs. J Am Coll Radiol 2022;19(2): 221–31. https://doi.org/10.1016/J.JACR.2021.09.047.
38 Heer E, Harper A, Escandor N, Sung H, McCormack V, Fidler-Benaoudia MM. Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study. Lancet Glob Health 2020;8(8):e1027–37. https://doi.org/10.1016/S2214-109X(20)30215-1.
39 Hoxha I, Islami DA, Uwizeye G, Forbes V, Chamberlin MD. Forty-five years of research and progress in breast cancer: progress for some, disparities for most. JCO Glob Oncol 2022;8(8). https://doi.org/10.1200/GO.21.00424.
40 Chen IY. Treating health disparities with artificial intelligence. Nat Med 2020;26: 16–7. https://doi.org/10.1038/s41591-019-0649-2.
41 Wilkinson MD, Dumontier M, Ijj Aalbersberg, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3(1):1–9. https://doi.org/10.1038/sdata.2016.18.
42 Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B. The citation advantage of linking publications to research data. PloS One 2020;15(4):e0230416. https://doi.org/10.1371/JOURNAL.PONE.0230416.
43. Vivli - Center for Global Clinical Research Data. https://vivli.org/. [Accessed 20 November 2023].
44. Home - Lacuna Fund. https://lacunafund.org/. [Accessed 20 November 2023].
45 Jeong JJ, Vey BL, Bhimireddy A, et al. The EMory BrEast imaging Dataset (EMBED): a racially diverse, granular dataset of 3.4 million screening and diagnostic mammographic images. Radiol Artif Intell 2023;5(1). https://doi.org/10.1148/RYAI.220047/ASSET/IMAGES/LARGE/RYAI.220047.TBL4.JPEG.
46 Frazer HML, Tang JSN, Elliott MS, et al. ADMANI: annotated digital mammograms and associated non-image datasets. Radiol Artif Intell 2023;5(2). https://doi.org/10.1148/RYAI.220072/ASSET/IMAGES/LARGE/RYAI.220072.FIG 3.JPEG.
47. USF Digital Mammography Home Page. http://www.eng.usf.edu/cvprg/Mammography/Database.html. [Accessed 16 August 2024].
48. DDSM Mammography. https://www.kaggle.com/datasets/skooch/ddsm-mammography. [Accessed 16 August 2024].
49 Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: toward a full-field digital mammographic database. Acad Radiol 2012;19(2):236–48. https://doi.org/10.1016/J.ACRA.2011.09.014.
50 Suckling J, et al. The Mammographic Image Analysis Society Digital Mammogram Database. International Congress Series 1069. 1994. p. 375–8.
51. Suckling J. The Mini-MIAS Database of Mammograms. http://peipa.essex.ac.uk/info/mias.html. [Accessed 20 November 2023].
52. Breast Cancer Digital Repository. https://bcdr.eu/. [Accessed 16 August 2024].
53 Halling-Brown MD, Warren LM, Ward D, et al. OPTIMAM mammography image database: a large-scale resource of mammography images and clinical data. Radiol Artif Intell 2021;3(1). https://doi.org/10.1148/RYAI.2020200103.
54. BREAST-CANCER-SCREENING-DBT - The Cancer Imaging Archive (TCIA). https://www.cancerimagingarchive.net/collection/breast-cancer-screening-dbt/. [Accessed 16 August 2024].
55. GitHub - Emory-HITI/EMBED_Open_Data: Data Descriptor and Sample Notebooks for the Emory Breast Imaging Dataset (EMBED) Hosted on the AWS Open Data Program. https://github.com/Emory-HITI/EMBED_Open_Data. [Accessed 12 August 2024].
56. CMMD - The Cancer Imaging Archive (TCIA). https://www.cancerimagingarchive.net/collection/cmmd/. [Accessed 12 August 2024].
57. Breast Micro-calcifications Dataset With Precisely Annotated Sequential Mammograms. doi:https://doi.org/10.5281/ZENODO.5036062.
58. CDD-CESM - The Cancer Imaging Archive (TCIA). https://www.cancerimagingarchive.net/collection/cdd-cesm/. [Accessed 16 August 2024].
59 Shin SY, Lee S, Yun ID, et al. A novel cascade classifier for automatic microcalcification detection. PloS One 2015;10(12). https://doi.org/10.1371/JOURNAL.PONE.0143725.
60. Digital mammography Dataset for Breast Cancer Diagnosis Research (DMID)DMID.rar. https://figshare.com/articles/dataset/_b_Digital_mammography_Dataset_for_Breast_Cancer_Diagnosis_Research_DMID_b_DMID_rar/24522883/1. [Accessed 12 August 2024].
61. VinDr-Mammo: a large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography v1.0.0. https://physionet.org/content/vindr-mammo/1.0.0/images/#files-panel. [Accessed 12 August 2024].
62. Mammographic Mass - UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/161/mammographic+mass. [Accessed 12 August 2024].
63. VICTRE - The Cancer Imaging Archive (TCIA). https://www.cancerimagingarchive.net/collection/victre/. [Accessed 12 August 2024].
64 Chen X, Zhang K, Abdoli N, et al. Transformers improve breast cancer diagnosis from unregistered multi-view mammograms. Diagnostics (Basel) 2022;12(7). https://doi.org/10.3390/DIAGNOSTICS12071549.
65. Ambra | Home. https://mamografiaunifesp.ambrahealth.com/. [Accessed 12 August 2024].
66. CSAW-S. doi:https://doi.org/10.5281/ZENODO.4030660.
67. Safe Haven (DaSH) | Research | The University of Aberdeen. https://www.abdn.ac.uk/research/digital-research/dash.php. [Accessed 12 August 2024].
68 Alsolami AS, Shalash W, Alsaggaf W, Ashoor S, Refaat H, Elmogy M. King Abdulaziz University Breast Cancer Mammogram Dataset (KAU-BCMD). Data 2021;6(11):111. https://doi.org/10.3390/DATA6110111.

69. Sinha GR. Mammograms-Breast Cancer Images. IEEE DataPort. doi:10.21227/9f0p-qx37.
70. RSNA Mammography Breast Cancer Detection PNG. https://www.kaggle.com/datasets/anitho2910/rsna-mammography-breast-cancer-detection-png/code. [Accessed 12 August 2024].
71. Full Mammogram. https://www.kaggle.com/datasets/princelubisi/full-mammogram. [Accessed 12 August 2024].
72 Aqdar KB, Abdalla PA, Mustafa RK, et al. Mammogram Mastery: A Robust Dataset for Breast Cancer Detection and Medical Education. 2024. p. 1. https://doi.org/10.17632/FVJHTSKG93.1.
73. BIRAD mammo. https://www.kaggle.com/datasets/sjudal/birad-mammo. [Accessed 12 August 2024].
74. Faramonna D. Breast Mammography Image Dataset With Masses. 2023. p. 1. https://doi.org/10.17632/8FZTXGGJNC.1.
75 Behravan H, Gudhe NR, Okuma H, Mannermaa A. Mammogram Density Assessment Dataset. 2023. p. 2. https://doi.org/10.17632/TDX3H2FN9V.2.