# Supplementary Information

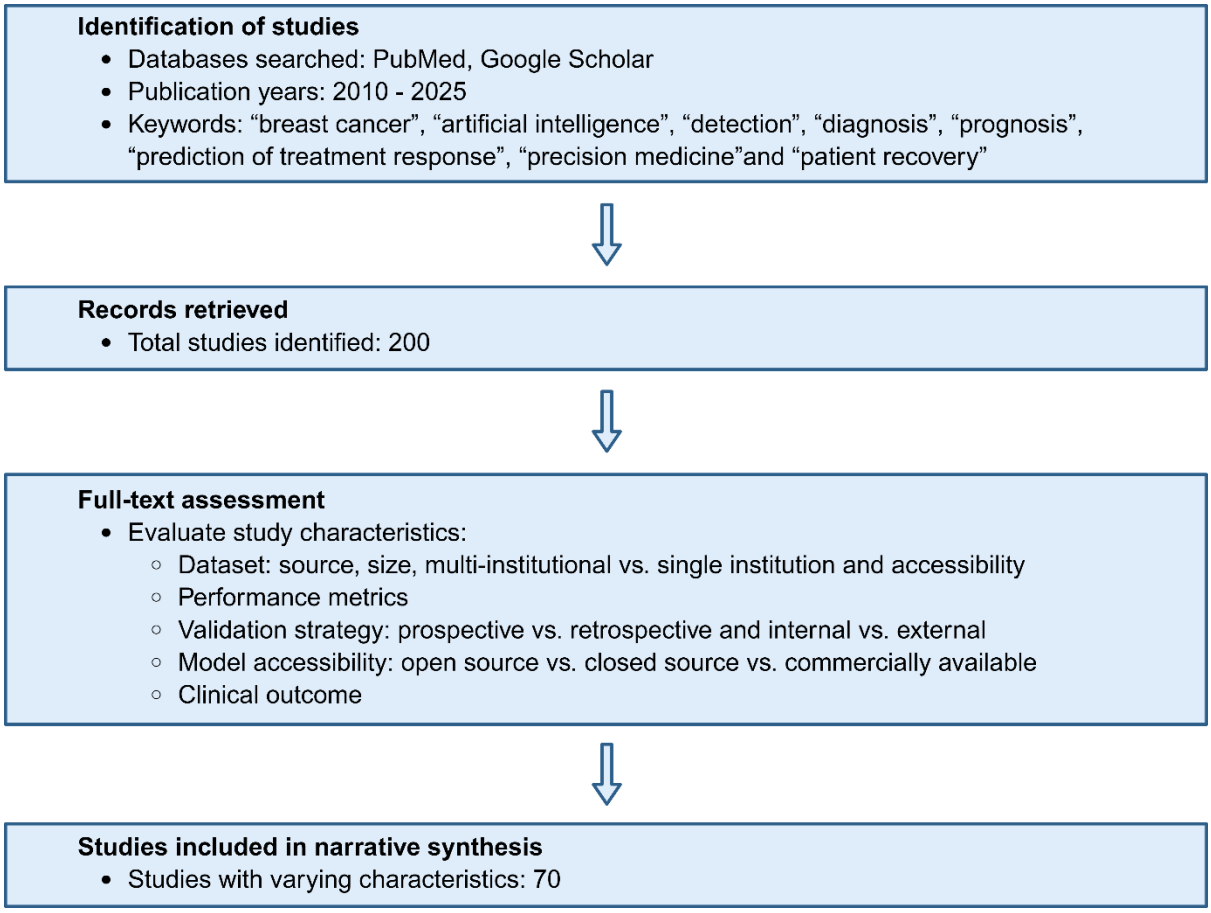
## Supplementary Methods

This review employed a narrative synthesis with a structured scoping approach rather than a formal systematic review. We searched databases including PubMed and Google Scholar for studies published between 2010 and 2025 that investigated the use of AI applications in breast cancer management. We searched for key terms such as “breast cancer”, “artificial intelligence”, “detection”, “diagnosis”, “prognosis”, “prediction of treatment response”, “precision medicine”, and “patient recovery”. Studies were not excluded on the basis of study design, validation strategy or dataset quality to provide a comprehensive overview of the strengths and weaknesses of the studies, as outlined in **Supplementary Figure 1**.

Studies that incorporated external validation, multi-institutional testing, or prospective designs provide stronger evidence of model robustness and clinical applicability compared to internal or retrospective-only validation. Cross-validation within a single dataset can demonstrate internal consistency, but it does not guarantee performance in diverse real-world settings, which raises concerns about generalisability to broader patient populations. External validation across independent cohorts, particularly those with demographic and institutional heterogeneity, offers a more reliable indication of generalisability. Prospective validation further strengthens confidence by assessing model performance in a clinical workflow before outcomes are known, thereby reducing bias and simulating real-world deployment.

Datasets in the selected studies ranged from small single-institution cohorts (fewer than 1,000 samples) to large multi-institutional repositories exceeding 100,000 cases. As such, the datasets varied greatly in terms of size and demographic diversity. This

variability introduces heterogeneity in reported performance metrics and limits comparability across studies.




**Supplementary Figure 1. Study selection process for artificial intelligence applications in breast cancer management.** The flowchart illustrates how the studies across the different phases of breast cancer management were selected and evaluated for this review. Included studies were synthesised to provide a comprehensive overview of methodological strengths, weaknesses and generalisability.

**Supplementary Table 1. Overview of studies evaluating AI applications in the pre-treatment phase – detection, diagnosis and prognosis, of breast cancer management.** The table summarises datasets (source, size, population-based vs. multi-institutional vs. single institution and accessibility), performance metrics, validation strategies (prospective vs. retrospective and internal vs. external), model accessibility (open source vs. closed source vs. commercially available), and reported clinical outcomes, providing a consolidated view of current evidence and translational potential. AR = androgen receptor, AUROC = area under the receiver operating characteristic curve, BCSC = breast cancer stem cell, ER = estrogen receptor, HER2 = human epidermal growth factor receptor 2, NPV = negative predictive value, PPV = positive predictive value, PR = progesterone receptor, TCGA = The Cancer Genome Atlas, TNBC = triple negative breast cancer and WSI = whole slide image.

| Purpose | Application | Dataset | Key Performance Metrics | Validation Strategy | Model Accessibility | Outcome | Reference |
|---|---|---|---|---|---|---|---|
| Detection | Mammogram | UK & US; ~28k women; multi-institutional; controlled access | AUROC, sensitivity & specificity | Retrospective & internal | Closed source | AUROC 0.871 vs. 0.750 (radiologists), ↑ sensitivity & specificity | [40] |
| | | Austria, Italy, Netherlands, Spain, Sweden, UK & US; ~189k mammograms; multi-institutional; controlled access | AUROC & sensitivity | Retrospective & internal | Commercially available: Transpara, ScreenPoint Medical | AUROC 0.840 vs. 0.814 (radiologists), ↑ sensitivity | [41] |
| | | Europe & US; ~18k mammograms; multi-institutional; controlled access | AUROC, sensitivity & specificity | Retrospective & internal | Commercially available: Transpara, ScreenPoint Medical | AUROC 0.89 vs. 0.87 (radiologists), sensitivity 86% vs. 83% & specificity 79% vs. 77% | [42] |
| | | South Korea, UK & US; ~170k exams; multi-institutional; controlled access | AUROC, sensitivity & specificity | Retrospective & external | Commercially available: Lunit INSIGHT MMG | AUROC 0.940 vs. 0.810 (radiologists), ↑ sensitivity & specificity | [43] |
| | | US; ~229k exams; single institution; controlled access | AUROC | Retrospective & internal | Open source | AUROC 0.876 vs. 0.705 - 0.860 (radiologists) | [44] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | BreastScreen Victoria, Australia; ~28k mammograms; population-based; controlled access | AUROC | Retrospective & internal | Closed source | AUROC 0.8979 | [45] |
| | | Germany; ~524k mammograms; population-based; controlled access | AUROC, sensitivity & specificity | Retrospective & external | Closed source | AUROC 0.982, sensitivity 89.8% vs. 87.2% (radiologists) & specificity 94.3% vs. 93.4% | [46] |
| | | Germany; ~2.3k women; population-based; non-image data available on request | Interval cancer detection rate | Retrospective | Commercially available: Vara, | Detected 12.2 - 27.5% of interval cancers | [47] |
| | | Norway; ~122k exams; population-based; controlled access | Interval cancer detection rate | Retrospective | Commercially available: Transpara, ScreenPoint Medical | Detected 30.7 - 44.9% of interval cancers | [48] |
| | | BreastScreen Western Australia; ~108k mammograms; | Interval cancer detection rate | Retrospective & external | Commercially available: Saige-Q, DeepHealth | Detected 36.6% of interval cancers | [49] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | population-based; controlled access | | | | | |
| | | Sweden, Mammography Screening with AI (MASAI) trial; ~80k women; population-based; available on request | Cancer detection rate & workload reduction | Prospective & randomised | Commercially available: Transpara, ScreenPoint Medical | Similar detection rate & 44.3% ↓ workload compared to standard double reading | [51] |
| | | Hungary; ~28k screens; single institution; available on request | Cancer detection rate & recall rates | Prospective | Commercially available: Mia, Kheiron Medical Technologies | 5 - 13% ↑ cancer detection rate & minimal to no unnecessary recalls | [52] |
| | | Sweden; ~55k women; population-based; available on request | Cancer detection rate | Prospective | Commercially available: Lunit INSIGHT MMG | 4% ↑ non-inferior detection compared to standard double reading | [53] |
| | | Germany; ~461k women; population-based; publicly available | Cancer detection rate & recall rate | Prospective | Commercially available: Vara | 17.6% ↑ cancer detection rate & no negative effect on recalls compared to standard double reading | [54] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Tomosynthesis | France, Germany & US; 80 cases; multi-institutional; controlled access | AUROC & reading time | Retrospective | Commercially available: SenoClaire, GE | 9.1s ↓ reading time & non-inferior AUROC | [61] |
| | | France & US; 240 cases; multi-institutional; controlled access | AUROC & reading time | Retrospective | Commercially available: PowerLook Tomo Detection, iCAD | 19s ↓ reading time & non-inferior AUROC | [62] |
| | | South Korea; 100 cases; single institution; controlled access | AUROC & reading time | Retrospective & internal | Closed source | 10.1s ↓ reading time & non-inferior AUROC | [63] |
| | | US; 260 exams; single institution; controlled access | AUROC & reading time | Retrospective | Commercially available: PowerLook Tomo Detection, iCAD | 33.7s ↓ reading time & 0.057 ↑ AUROC | [64] |
| | | US; 240 cases; single institution; cases from NCT01373671 | AUROC & reading time | Retrospective | Commercially available: Transpara, ScreenPoint Medical | 5s ↓ reading time & 0.03 ↑ AUROC | [65] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | MRI | US & Germany; 111 women; multi-institutional; controlled access | AUROC & sensitivity | Retrospective & external | Closed source | AUROC ↑ from 0.71 to 0.76 & ↑ sensitivity than radiologists | [72] |
| | | Portugal; 124 patients; single institution; controlled access | AUROC | Retrospective & internal | Closed source | Comparable AUROC 0.778 vs. other studies | [73] |
| | | US & Poland; ~13k patients; multi-institutional; controlled access | AUROC & biopsy reduction | Retrospective & external | Open source | AUROC 0.924 vs. 0.890 (radiologists) & ↓ unnecessary biopsies | [74] |
| | | Netherlands; 361 women; single institution; controlled access | Sensitivity | Retrospective | Closed source | ↑ sensitivity than radiologists & automated process | [75] |
| | | Korean; 101 women; single institution; controlled access | Image realism | Retrospective & internal | Open source | Simulated images indistinguishable from real scans & eliminated need for contrast agent | [76] |
| | Ultrasound | Egypt; 600 women; single institution; publicly available | Accuracy | Retrospective & internal | Closed source | 97.18% classification accuracy (normal/benign/malignant) | [78] |
| | | China; ~1.3k women; multi- | AUROC, sensitivity & specificity | Prospective & internal | Open source | Comparable AUROC 0.902, sensitivity 75.2% & | [79] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | institutional; controlled access | | | | specificity 91.8% vs. radiologists | |
| | | China; ~6.7k women; multi-institutional; controlled access | AUROC | Retrospective & external | Open source | AUROC 0.945 vs. 0.716 (radiologists) & ↓ unnecessary biopsies | [80] |
| | | US; ~1.3k images; single institution; controlled access | Sensitivity, PPV & inter-operator variability | Retrospective | Commercially available: Koios DS, Koios Medical | ↑ sensitivity, PPV & ↓ variability vs. radiologists | [81] |
| Diagnosis | Genomic and transcriptomic profiling | TCGA; ~1.2k tumour samples; multi-institutional; publicly available | Accuracy | Retrospective & internal | Closed source | ≤86% accuracy distinguishing TNBC vs. non-TNBC | [83] |
| | Histopathology | US; ~1.2k patients; single institution; available on request | Accuracy | Retrospective & internal | Available on request | Accuracy 75% in predicting grade, ER status, subtype & recurrence | [87] |
| | | South Korea; 401 patients; single institution; controlled access | AUROC | Retrospective & internal | Open source | AUROC 0.75 - 0.91 vs. 0.67 - 0.83 (2D models) in predicting ER, PR, AR, HER2 status and Ki-67 index | [88] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | US; ~6k nuclei from 7 patients; single institution; controlled access | Labour & time | Retrospective & internal | Closed source | ↓ labour and time for HER2 amplification detection | [89] |
| | Nanotechnology / Nanomedicine | Taiwan; 439 samples; single institution; available on request | Accuracy, AUROC, PPV, NPV, sensitivity & specificity | Retrospective & internal | Closed source | Accuracy 91%, AUROC 0.99, PPV 97%, NPV 97%, sensitivity 86% & specificity 97% (electric nose) | [91] |
| | | US; 120 samples from 5 breast cell lines; single institution; controlled access | Accuracy | Retrospective & internal | Closed source | Accuracy 98.1% (carbon nanoparticle platform) | [92] |
| Prognosis | Histopathology | US; 67 patients; single institution; available on request | AUROC | Retrospective & internal | Closed source | AUROC 0.57 - 0.68 predicting Oncotype DX | [95] |
| | | France & Netherlands; 174 WSIs; multi-institutional; available on request | Accuracy | Retrospective & internal | Closed source | 83.19% accuracy predicting Oncotype DX | [96] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | US; 560 samples from 122 patients; multi-institutional; publicly available | Prognostic correlation | Retrospective & internal | Open source | Cell-state proportion and organisation were predictive of disease stage | [97] |
| | | Canada; 328 women; multi-institutional; publicly available | Accuracy, sensitivity & specificity | Retrospective & internal | Open source | 97% accuracy, 98% sensitivity, 80% specificity predicting Oncotype DX risk score | [98] |
| | | US; 41 images from 12 patients; single institution; controlled access | Accuracy | Retrospective & internal | Closed source | 90% accuracy grading lymphocytic infiltration in HER2+ breast cancer | [99] |
| | Genomic, transcriptomic and metabolomic data | TCGA; 1069 patients; multi-institutional; publicly available | AUROC | Retrospective & external | Available on request | BCSC risk score AUROC 0.746 - 0.805 for predicting survival & CD79A+CD24-PANCK+- BCSCs associated with poor prognosis | [100] |
| | | China; 330 TNBC and 149 paired normal samples; single institution; publicly available | Recurrence risk stratification | Retrospective & external | Closed source | Stratified TNBC into high/low recurrence groups | [101] |

**Supplementary Table 2**. **Overview of studies evaluating AI applications in the treatment phase – prediction of treatment response and precision medicine, of breast cancer management.** The table summarises datasets (source, size, multi-institutional vs. single institution and accessibility), performance metrics, validation strategies (prospective vs. retrospective and internal vs. external), model accessibility (open source vs. closed source vs. commercially available), and reported clinical outcomes, providing a consolidated view of current evidence and highlighting the potential of AI to guide therapeutic decision-making. AUROC = area under the receiver operating characteristic curve, CDK4/6 = cyclin-dependent kinase 4/6, CTRP = Cancer Therapeutics Response Portal, GDSC = Genomics of Drug Sensitivity in Cancer, HR = hazard ratio, $IC_{50}$ = half maximal inhibitory concentration, MSE = mean squared error and TCGA = The Cancer Genome Atlas.

| Purpose | Application | Dataset | Key Performance Metrics | Validation Strategy | Model Accessibility | Outcome | Reference |
|---|---|---|---|---|---|---|---|
| Prediction of Treatment Response | Predict pathologic complete response to neoadjuvant therapy | UK; 147 cases; multi-institutional; publicly available | AUROC | Retrospective & external | Open source | AUROC 0.87 | [104] |
| | | China; 297 patients; multi-institutional; controlled access | AUROC | Retrospective & external | Closed source | AUROC 0.896 - 0.903 | [105] |
| | | US; 81 tumours from 80 patients; single institution; controlled access | AUROC | Retrospective & internal | Closed source | AUROC 0.91 | [106] |
| | Predict survival & $IC_{50}$ drug response | TCGA & GDSC; 532 patients & 42 cell lines; multi-institutional; publicly available | AUROC, MSE & overall regression value | Retrospective & external | Open source | AUROC 0.98 (survival); MSE 1.154 mean square error & overall regression value 0.92 ($IC_{50}$) | [107] |
| | Predict sensitivity to CDK4/6 inhibitor | CTRP & GDSC; 1244 cell lines; multi-institutional; publicly available | HR | Retrospective & internal | Open source | HR 0.21 (sensitive vs. resistant) | [108] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Predict cardiotoxicity in response to anthracycline treatment | Taiwan; 211 patients; single institution; controlled access | AUROC | Prospective & internal | Closed source | AUROC 0.66 (cardiac dysfunction) & 0.81 (heart failure) | [109] |
| Precision Medicine | Drug Combination Prediction and Testing (DCPT) platform | Finland; 3 patients; single institution; available on request | Personalised drug combination | Prospective | Open source | Identified effective drug combinations for T-cell prolymphocytic leukemia patients | [127] |
| | Quadratic Phenotypic Optimization Platform (QPOP) | Singapore; 18 patient-derived avatars; multi-institutional; available on request | Personalised drug combination | Prospective | Closed source | Identified effective drug combinations for hepatocellular carcinoma | [128] |
| | | Singapore; 13 patients; single institution; publicly available | Personalised drug combination | Prospective | Closed source | Identified effective drug combinations for multiple myeloma | [129] |
| | | Singapore; 1 patient; single | Personalised drug combination | Prospective | Closed source | Identified effective drug | [130] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | institution; publicly available | | | | combinations for refractory T-cell lymphoma | |
| | | Singapore; 75 samples from 71 patients; multi-institutional; publicly available | Personalised drug combination | Prospective | Closed source | Identified effective drug combinations for refractory lymphoma | [131] |
| | CURATE.AI | Singapore; 1 patient; single institution; publicly available | Personalised drug dosing | Prospective | Closed source | Optimised combination drug dose for metastatic prostate cancer | [133] |
| | Automated genetic counselling chatbot | US; 37 patients; single institution; controlled access | Personalised counselling | Prospective & randomised | Commercially available: Gia, Invitae | Performed on par with genetic counsellor in terms of patient satisfaction and comprehension | [135] |

97