

1           **Present and Future of Machine Learning in Breast Surgery: a systematic review**

2                           **(Running Head: Machine Learning in Breast Surgery)**

3  
4   **AUTHORS:**

5   Chien Lin Soh<sup>1\*</sup>; Viraj Shah<sup>2\*</sup>; Arian Arjomandi Rad<sup>2,3</sup>; Robert Vardanyan<sup>2</sup>; Alina Zubarevich<sup>4</sup>,  
6   Saeed Torabi<sup>5</sup>, Alexander Weymann<sup>4</sup>, George Miller<sup>3,6</sup>; Johann Malawana<sup>3,6</sup>.

7   **INSTITUTION:**

- 8           1. School of Clinical Medicine, University of Cambridge, Cambridge, UK.
- 9           2. Department of Medicine, Faculty of Medicine, Imperial College London, London,  
10           United Kingdom.
- 11           3. Research Unit, The Healthcare Leadership Academy, London, United Kingdom.
- 12           4. Department of Thoracic and Cardiovascular Surgery, West German Heart and  
13           Vascular Center Essen, University Hospital of Essen, University Duisburg-Essen,  
14           Essen, Germany.
- 15           5. Department of Anesthesiology and Intensive Care Medicine, University Hospital of  
16           Cologne, Cologne, Germany
- 17           6. Centre for Digital Health and Education Research (CoDHER), University of Central  
18           Lancashire Medical School, Preston, United Kingdom.

19   **\*Authors contributed equally**

20   **Corresponding author:** Arian Arjomandi Rad, Imperial College London, Department of  
21   Medicine, Faculty of Medicine, South Kensington Campus, Sir Alexander Fleming Building,  
22   London, United Kingdom. Email: [arian.arjomandi-rad16@imperial.ac.uk](mailto:arian.arjomandi-rad16@imperial.ac.uk)

23 **Word count: 5374**

24 **Number of Figures: 3**

25 **Number of Tables: 1**

26 **Conflict of interest: none to declare**

27 **Funding: none to declare**

28 **Data availability: data collection form and search results are available on enquiry to the**

29 **corresponding author (A.AR)**

30

31

32

33

34

35

36

37

38

39

40

41

42

43 **Abstract:**

44 **Background:** Machine learning (ML) is a set of models and methods that can automatically detect  
45 patterns in vast amounts of data, extract information and use it to perform decision making under  
46 uncertain conditions. The potential of ML is significant, and breast surgeons must strive to be informed  
47 with the up-to-date knowledge and its applications. Here, we aim to review the current applications  
48 of ML in breast surgery.

49 **Methods:** A systematic database search was conducted of original articles that explored the use of ML  
50 and/or AI in breast surgery in EMBASE, MEDLINE, Cochrane database and Google Scholar, from  
51 inception to December 2021.

52 **Results:** Our search yielded 477 articles, of which 14 studies were included in this review, featuring  
53 73,847 patients. Four main areas of application were identified: 1) ML for predictive modelling of  
54 breast surgical outcomes; 2) ML in breast image-based context for analysis and detection, including  
55 mammography; 3) ML within screening and triaging of breast surgery patients; 4) ML network utility  
56 for detective purposes. There is evident value to the use of ML in pre-operative planning and provision  
57 of information for breast surgery in a cancer and an aesthetic context. ML outperformed traditional  
58 statistical modelling in all studies for predicting mortality, morbidity, and quality of life outcomes. ML  
59 patterns and associations could support planning, anatomical visualisation, and surgical navigation.

60 **Conclusion:** ML demonstrated promising applications for improving breast surgery outcomes and  
61 patient-centred care, nevertheless, there remain important limitations and ethical concerns relating  
62 to implementing AI into everyday surgical practices.

63

64

65

66

67 **Background:**

68 Artificial intelligence (AI) refers to computer systems that mimic human cognitive functions and learn  
69 using large data sets. (1) Recent years have shown a dramatic development in these technologies in  
70 healthcare employed in a wide variety of diagnostic and decision-making processes. (2) In an emerging  
71 era of big data, the scope and scale of patient data available and leaps in computational ability has  
72 allowed AI to develop and improve in its efficiency and applicability. (3)

73 AI technology is progressing rapidly with support from healthcare professionals, industry and  
74 governments. (4) Healthcare has adopted these technologies to improve patient outcomes, especially  
75 in the field of surgery. These technologies demonstrate unique potential in breast surgery: pre-  
76 operative planning, patient outcome predictions, and even overcoming the challenges of the COVID-  
77 19 pandemic as demonstrated by the recent COVIDSurg Collaborative study that addressed the impact  
78 of COVID-19 on patient mortality with a predictive model. (5)

79 AI encompasses many disciplines of computer learning, and clinically relevant subtypes of AI include  
80 machine learning (ML). (1,6) Machine learning is a subset of this field where a system focuses on using  
81 algorithmic packages and data to mimic the way humans learn. (2) The algorithms use data inputs to  
82 'learn', uncovering associations in data sets via pattern recognition, repetition and modification to  
83 make autonomous decisions and predict future outcomes. Common subsets of ML include prediction  
84 models, deep learning and natural language processing. (7-8)

85 Breast surgery, a sub-speciality within general surgery, is a field that has much to benefit from the  
86 advances in AI to provide the best patient care by surgical interventions in benign and malignant breast  
87 disease. ML in breast surgery may involve these sets of models and methods to detect patterns in vast  
88 amounts of patient data, extract appropriate information and use it to perform decision making under  
89 uncertain conditions. (9) From supporting pre-operative planning to predicting future outcomes of  
90 surgery. The potential applications of ML is significant, and breast surgeons must strive to be informed  
91 with the up-to-date knowledge and applications of this subset of AI within the speciality. (10-11)

92 The aim of this review is to study the applications of ML in breast surgery. Past reviews in other surgical  
93 specialities have been written, but none specifically for breast surgery. This review is designed to  
94 closely evaluate the current applications by synthesising current research, and to catalyse future  
95 research efforts into this advancing field.

96

## 97 **Methods**

### 98 **Literature Search Strategy**

99 This systematic review was conducted in accordance with the Cochrane Collaboration and Preferred  
100 Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A literature search  
101 was conducted including EMBASE, MEDLINE, Cochrane, PubMed and Google Scholar from inception  
102 to December 2021 (*Figure 1*). The search terms used were (Machine Learning OR Artificial Intelligence  
103 OR Deep learning OR Decision Trees OR Neural Networks) AND (Breast Surgery OR Mastectomy OR  
104 Breast-conservative Surgery OR Breast reduction OR Breast reconstruction OR Breast augmentation  
105 OR Breast Cancer Surgery). Further articles were identified through use of the ‘related articles’  
106 function on MEDLINE and a manual search of the references lists of articles found through the original  
107 search. The only limits used were the English language and the aforementioned time frame.

### 108 **Study inclusion and exclusion criteria**

109 All original articles were included reporting the use of machine learning in breast surgery. Studies were  
110 considered if they presented ML models with the aim of supporting breast surgery or providing a  
111 prognosis for an intervention, either used by itself or with other methods. There were no geographical  
112 restrictions. Studies were excluded from the review if the quality of available data and data  
113 inconsistencies precluded valid extraction or if the study was performed in an animal model. Case  
114 reports, reviews, abstracts from meetings and preclinical studies were excluded. Machine learning is  
115 a highly erratic and dynamic field – this review contains literature published over a 5-year time period

116 between 2017 and 2021 inclusive, technology has changed significantly even in the five years  
117 preceding conduction of this review. As a result, there have been many advancements that have  
118 superseded some of the points raised in earlier literature and care was taken to recognise each study  
119 in the unique context of its publication year. It was ensured that any outdated findings did not shape  
120 the review. By following the aforementioned criteria, two reviewers (C.S and V.S.) independently  
121 identified articles for further assessment following title and abstract review. Disagreements between  
122 the two reviewers were resolved by a third independent reviewer (A.AR.). Potentially eligible studies  
123 were then retrieved for full text assessment.

#### 124 **Data extraction and critical appraisal of evidence**

125 All full texts of retrieved articles were read and reviewed by two authors (C.S. and V.S.) and inclusion  
126 or exclusion of studies was decided unanimously. When there was disagreement, a third reviewer  
127 (A.AR.) made the final decision. Using a pre-established protocol, the following data was extracted:  
128 first author, study type and characteristics, number of patients, population demographics, Type of  
129 Procedure, Category of machine learning method utilised, Method of machine learning implemented  
130 and Main reported outcomes.

#### 131 **Risk of Bias**

132 The risk of bias of the selected articles were evaluated by two independent reviewers (C.S. and V.S.)  
133 using an adapted Cochrane Collaboration Risk of Bias tool (*Figure 2*). The methodological quality of  
134 the studies were assessed based of domains: 1. Study Participation, 2. Study Response, 3. Outcome  
135 Measurement, 4. Statistical Analysis and Reporting, 5. Study Confounding. An overall grading of low,  
136 medium or high risk of bias were then allocated.

137 Additionally, the limitations of this systematic review have been more expansively outlined in  
138 [Supplementary File 1](#).

139

140 **Results:**

141 **Study selection**

142 The literature search identified 477 articles, of which 361 were screened following removal of  
143 deduplicates and 24 were full text reviewed and assessed in accordance with the inclusion and  
144 exclusion criteria. Following critical appraisal, a total of 14 studies (13-26) were included in this review,  
145 featuring 73,847 patients. Figure 1 illustrates the entire study selection process. A summary of the  
146 studies collected and their respective designs, type of machine learning mode used and its  
147 implementation as well as the main reported outcomes are found in table 1.

148 There were 9 studies (16-18,20-23,25,26) (5–7,9–12,14,15) which described examples of machine-  
149 learning based predictive modelling comprising 45792 patients and included a conglomerate of  
150 different modelling methods. Predictive modelling was the use with the most recorded studies and  
151 patient volume. The use of ML in imaging is also described. There were 3 studies (13,15,24) (2,4,13)  
152 which described examples of machine-learning within an image-based context for analysis and  
153 detection comprising 20499 patients – in all cases different modes of machine learning modes were  
154 applied. There was a one study (14) (3) which described a case of machine learning’s role within  
155 screening and triaging – this comprised 7364 patients. There was one study (19) which described a  
156 scenario of machine learning network utility for detective purposes – this comprised 355 patients.

157 (8)

158 **Challenges and recommendations**

159 *Figure 3* summarizes the main challenges and respective recommendations developed from the  
160 literature with regards to future research being conducted in the field of ML and its application in  
161 breast surgery. The recommendations should be taken into the context of each future research study  
162 and be considered with all the information available.

163

164 **Discussion:**

165 This systematic review provides a wide-scope summary of the uses of machine learning and artificial  
166 intelligence within breast surgery. Although the review's results demonstrate successes within  
167 different approaches in the field, these must be considered in the context of their limitations and  
168 recent applications. Most of these applications remain at the 'proof of concept' stage. Healthcare  
169 professionals must be prepared to adopt machine learning and artificial intelligence into health  
170 practice, as usability and cost-effectiveness of these technologies increase with new developments in  
171 the field, and to shape the new landscape in which it is used in medicine. (11)

172 There is evident value to the use of machine learning in pre-operative planning and provision of  
173 information for breast surgery planning in a cancer and an aesthetic context. The diagnosis and  
174 detection of pathology is fundamental in pre-operative planning for breast cancer resection. The use  
175 of image analysis in clinical applications such as breast imaging, digital pathology and surgical planning  
176 has been well-described in the literature. Considering the imaging data available with modern  
177 radiology techniques in screening and diagnosis, many machine learning solutions have been derived.

178 A retrospective image analysis study from Becker et al (13) (2) describes mammography diagnostics  
179 using neural network image analysis software that demonstrated the equivalent performance of the  
180 neural network with an area under the receiver operating characteristic curve (AUC) of 0.81 and 0.82  
181 respectively in both stages compared to radiologists. The results differ between radiologists - however,  
182 the neural network showed an increased sensitivity of 72% when compared to a 66.7% average across  
183 the radiologists overall. Corroborating these findings, the retrospective simulation study from  
184 Dembrower et al (14) (3) and diagnostic study from Buda et al (15)(4) also demonstrate similar levels  
185 of prediction using different means of deep learning models such as a commercial artificial  
186 intelligence-based cancer-detector algorithm. Cancer detection via datasets allows specific  
187 interventions that increase efficiency of operations with the pre-knowledge for surgeons to employ in  
188 planning and treatment decisions. The employment of specific machine learning technologies

189 including the Faster-RCNN with Inception-ResNet-v2 deep-learning framework, described in the Yap  
190 et al study for a set of ultrasound breast images, facilitates the accurate detection of objects to allow  
191 surgeons to focus on the relevant area of the breast (24) (13).

192 The models that exist can promise a future of reducing the numbers of biopsies and bringing efficiency  
193 to radiologist interpretation while reducing workloads. Diagnostic and prognostic applications in  
194 imaging and pathology have been studied greatly, with a wide evidence base of applied research.  
195 Transfer of imaging information to the operating theatre to more accurately localise cancer to support  
196 surgical care can aid the field. Reducing overdiagnosis, morbidity and time efficiency with the efficient  
197 use of breast imaging is an aspiration of the future.

198 Most applications of machine learning within breast surgery, as ascertained by this review, centre  
199 around the prediction of patient outcomes prognostically – this is coherent with the wider applications  
200 of machine learning within modern surgery (27-29)(16–18). There have been multiple instances,  
201 identified across the course of this review, where machine learning was employed alongside  
202 traditional statistical modelling for predictive purposes. Indicative of the success of machine learning,  
203 the former outperformed the latter in every highlighted example and thus replicates the successes  
204 seen across other surgical subspecialties, mostly prominently neurosurgery (30) (19). This is  
205 particularly evidenced via the longitudinally-designed Huang et al study which demonstrated that 5-  
206 year mortality after breast cancer surgery could be accurately assessed using machine learning  
207 algorithms through a variety of input variables (16) (5). Although machine learning packages have  
208 been substantiated to show marked improvements to pre-existing models including, but not limited  
209 to, least square regression and Cox regression (17,31) (6,20), it must be stated that these still remain  
210 limited and relatively novel.

211 The studies included in this review demonstrated high heterogeneity in the form of machine learning  
212 applied within individual cases. In this way, some models have been transparently more consistent  
213 and accurate than others. Artificial neural networks, algorithms which have been modelled after the

214 human brain and nervous system and with a basis in the concept of a human neurone (32) (21), have  
215 been established as dependable across the scope of this review – exemplified most prominently in the  
216 Lou et al. study where the artificial neural network package demonstrated the highest prediction  
217 performance index - obtaining a sensitivity of 95.9, specificity of 99.5, PPV of 99.0, NPV of 99.1 and  
218 AUC of 97.6 (20) (9). This provides greater support to previous literature which describes the  
219 effectiveness of artificial neural networks in other clinical contexts (33-35) (22–24). Artificial neural  
220 networks are better adapted to deal with more problematic inputs - specifically, cases where an input  
221 may be noisy or incomplete. As a result of these advantages, a 93.75% accuracy rate of identifying a  
222 breast cancer patient’s post-operative lymphedema status was obtained in the Fu et al. Study (19)(8).  
223 Many medical databases, of the scale where a machine learning model can be realistically derived,  
224 contain non-normally distributed data. This poses further issue to many forms of modelling which  
225 assume a normal distribution within a dataset (23) (12) - as artificial neural networks are applicable to  
226 well-correlated data that are not necessarily natively normally distributed, artificial neural networks  
227 are more transferrable and can provide greater potential for use in wider treatment contexts beyond  
228 breast cancer surgery.

229 Machine learning’s capacity for use in breast surgery can further extend past predicting outcomes and  
230 pivot towards providing more holistic patient assessment via the prediction of postsurgical pain, seen  
231 in both the Juwara et al (17)(6) and Lotsch et al (18) (7) studies. Machine learning creates opportunities  
232 for far more efficient pain assessment that can be undertaken immediately post-operatively, in  
233 comparison to pre-existing tools that require time-heavy questionnaires and extensive clinician-  
234 patient interaction; this serves great utility in the context of a healthcare system where both time and  
235 human resources are often limiting factors. As neuropathic pain can be debilitating for patients, early  
236 prediction can allow clinicians to better optimise post-surgical care.

237 As an additional factor repeatedly indicated via the machine learning models included in this review,  
238 a greater surgeon volume was the largest predictor of reduced breast cancer recurrence in patients

239 that had undergone breast cancer surgery (16,20) (5,9). In light of this, there is renewed necessity for  
240 machine learning to be implemented to further review the decision-making of surgeons with higher  
241 operative volumes. Decision analysis and reinforcement learning, modes of machine learning well-  
242 documented within this context (36) (25), would allow this to be paralleled and promote further  
243 improvements in decision-making for surgeons with lower operative volumes. This can further  
244 minimise postoperative burden. In tandem to this, there has been evidence to suggest that some  
245 machine learning packages can outperform even the most experienced surgeons, as determined  
246 above, and therefore it may be possible to provide the framework of a specific machine learning  
247 package itself as a template for replication by surgeons of all grades.

248 Additional applications of machine learning on the horizon within the field of breast surgical care can  
249 be considered, although most of these remain conceptual. Decision making in modern medicine is  
250 complex due to the increasing availability of data to consider before treatment decisions are made  
251 (37)(26). Advances in medical knowledge including that of well-researched novel therapies and  
252 surgery only dramatically increase the potential treatment choice algorithms. Decision support  
253 systems have been well-described, including the DESIREE project (38)(27), that provides breast unit  
254 physicians with decision support modules. This is further stipulated in other examples (39) (28) where  
255 decision support models regarding recurrence prediction are employed and comprise support systems  
256 that encompass artificial intelligence and information visualisation amongst many other technologies.

257 Computer vision for object and scene recognition could support surgical techniques. Patterns and  
258 associations can support planning, anatomical visualisation and surgical navigation. The exploration  
259 of machine learning systems that perform or directly complement surgery is rapidly developing, and  
260 may be a possibility in the imminent future. Real-time decision making supported by machine learning  
261 provides exciting opportunities (40) (29).

262 The new frontier of surgical innovation, concurrently occurring at the time of the review, is unlike any  
263 other observed previously. Despite the benefits and applications in the field as described above,

264 clinicians must be careful to consider the potential limitations and risks of the technology and avoid  
265 overt optimism surrounding the capabilities of machine learning, instead focusing realistically on the  
266 barriers to its implementation clinically. (41) (30).

267 Machine learning and artificial intelligence is limited by the lack of accurate and unbiased data  
268 collection and input. If data-input bias is evident, predictions may easily become unreliable. Examples  
269 include systematic biases due to nonrepresentative predictions for patient groups that are not  
270 necessarily represented in research (42) (31). This review does provide evidence in support of  
271 theoretical machine learning applications, however, as outlined in a review by Manlhiot et al. (43),  
272 care must be taken to recognise that these are not wholly clinically representative. The described  
273 machine learning models rely on heavily-curated datasets with relatively few implementation  
274 obstacles, in vast contrast to datasets available in clinical practice.

275 Machine learning can exhibit 'black box' characteristics, with incomprehensibly complex algorithms  
276 for their outputs. The learning mechanisms of some machines have been difficult to reproduce, and it  
277 has been difficult to justify certain decisions. Measures taken in the programming of these machines  
278 to justify decisions and compare with gold standard diagnosis can circumvent this challenge. The  
279 challenges surrounding the complexity of machine learning in its current state render it clinically  
280 unimplementable without expertise and specialist knowledge - hence, the benefits of it are yet to be  
281 properly actioned. Explainable machine learning whereby the system is able to justify how it made its  
282 predictions on a level that is comprehensible to a clinician(44,45) (32,33) has come to light as a  
283 potential solution to this issue - exemplified in the Moncada-Torres et al study (25) (14) where  
284 explainable machine learning consistently outperformed all other methods for predicting outcomes  
285 in patients who had undergone breast-conserving surgery or mastectomy. However, this remains one  
286 of a set of isolated cases.

287 In addition, considerations of collaboration with other stakeholders in the implementation of the  
288 technology, in order to ensure data is interpreted correctly and applied in the correct manner, should

289 be of paramount importance. These technologies have been criticised due to a lack of confidence  
290 regarding their perceived privacy and due to the risk of bias with data collection, as aforementioned  
291 Caution and planning to understand the most safe and beneficial method of implementation, via close  
292 collaboration of healthcare professionals with machine learning and artificial intelligence experts in a  
293 multidisciplinary approach, is required to ensure the best outcomes for all. In addition, engagement  
294 of breast surgery patients in decisions where patients can be informed are important.

295 Economic considerations, job losses and the lack of human element pose additional ethical dilemmas.  
296 Machine learning may be stifled from practical implementation in breast surgery due to infrastructural  
297 shortcomings (with regards to both hardware and software) in the post-deployment management of  
298 models, a phenomenon that has been described within the technology's application within cardiology  
299 (43). Ethico-legal and social issues including the lack of regulatory structures surrounding the  
300 technology must be addressed and solutions to such issues within breast surgical care must be  
301 explored. Financial considerations and the accessibility of this technology in LMICs and globally,  
302 alongside the opportunity-cost of such applications, must also be considered. The technology should  
303 be widely accessible, and it is important to ensure the maintenance of high standards across different  
304 healthcare systems.

305 The most favourable studies included in this review included high sample sizes and were multicentre.  
306 Many of the single centre studies cited the idea of applying the relevant algorithm to larger sample  
307 sizes through involving data from other centres. Additionally, many studies circumnavigated the  
308 challenge of a low centre sample size by combinations with registry data to build their respective  
309 algorithms. Potential prospective solutions may also have basis in the concept of federated learning –  
310 a machine learning approach that allows an algorithm to combine data collectively from multiple  
311 centres without physical exchange of the data (46). Hence, it is clear that any future approaches should  
312 ensure that this collaborative approach is undertaken as standard. Many studies encountered  
313 additional issues with data imbalances such that some classes contained significantly larger amounts

314 of data compared to others – to correct this, the Myung et al. study applied the ROSE:sMOTE  
315 oversampling technique which had been seen in previous literature (47) (34) and is a technique future  
316 studies may consider employing to increase the validity and generalisability of packages and  
317 consequentially the probability of success (21)(10).

318 Machine learning must be recognised as still within its trial phase – it is not perfect and subject to  
319 multiple flaws (26) (15). Although current literature provides fundamental foundations to its  
320 applicability, future approaches must consider clinical relevance at their core. The exact framework  
321 by which machine learning may seek to aid postoperative care must be established, as does the basis  
322 of a clinical decision support system. Such a system would aid the facilitation of greater data-based  
323 shared patient-clinician decision-making within breast surgical care. Hence, there is sufficient  
324 groundwork to construct prospective randomised studies to observe the impact of machine learning  
325 in clinical practice.

326 Significant and structured research is required to investigate the accuracy and utility of machine  
327 learning for the benefit of patients. It is important for surgeons to collaborate with computer scientists  
328 and related field specialists, so outcomes are directed towards improving patient care and to  
329 understand important limitations and ethical concerns relating to implementing such technology into  
330 everyday surgical practices.

331

332

333

334

335

336

337 **References:**

- 338 1. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive  
339 Computing: What Do These Terms Mean and How Will They Impact Health Care? J  
340 Arthroplasty. 2018;33(8):2358-61.
- 341 2. Topol EJ. High-performance medicine: the convergence of human and artificial  
342 intelligence. *Nature Medicine*. 2019;25(1):44-56.
- 343 3. Helm JM, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, et al. Machine  
344 Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Curr*  
345 *Rev Musculoskelet Med*. 2020;13(1):69-76.
- 346 4. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems.  
347 *Journal of global health*. 2018;8(2):020303-.
- 348 5. Machine learning risk prediction of mortality for patients undergoing surgery with  
349 perioperative SARS-CoV-2: the COVIDSurg mortality score. *Br J Surg*. 2021;108(11):1274-  
350 92.
- 351 6. Jarvis T, Thornburg D, Rebecca AM, Teven CM. Artificial Intelligence in Plastic Surgery:  
352 Current Applications, Future Directions, and Ethical Implications. *Plastic and*  
353 *reconstructive surgery Global open*. 2020;8(10):e3200-e.
- 354 7. Naylor CD. On the Prospects for a (Deep) Learning Health Care System. *JAMA*.  
355 2018;320(11):1099-100.
- 356 8. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine.  
357 *J Family Med Prim Care*. 2019;8(7):2328-31.
- 358 9. Ferreira MF, Savoy JN, Markey MK. Teaching cross-cultural design thinking for  
359 healthcare. *Breast*. 2020;50:1-10.
- 360 10. Coiera E. The Price of Artificial Intelligence. *Yearb Med Inform*. 2019;28(1):14-5.
- 361 11. Tang A, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J, et al. Canadian  
362 Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Can Assoc*  
363 *Radiol J*. 2018;69(2):120-35.
- 364 12. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial  
365 intelligence in healthcare: a structured literature review. *BMC Medical Informatics and*  
366 *Decision Making [Internet]*. 2021 [cited 31 December 2021];21(1). Available from:  
367 [https://bmcmidinformedecismak.biomedcentral.com/articles/10.1186/s12911-021-](https://bmcmidinformedecismak.biomedcentral.com/articles/10.1186/s12911-021-01488-9)  
368 [01488-9](https://bmcmidinformedecismak.biomedcentral.com/articles/10.1186/s12911-021-01488-9)
- 369 13. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in  
370 mammography diagnostic accuracy of a multipurpose image analysis software in the  
371 detection of breast cancer. *Investigative Radiology [Internet]*. 2017 [cited 2021 Dec  
372 3];52(7):434–40. Available  
373 from: [https://journals.lww.com/investigativeradiology/Fulltext/2017/07000/Deep\\_Lear](https://journals.lww.com/investigativeradiology/Fulltext/2017/07000/Deep_Learning_in_Mammography_Diagnostic_Accuracy.7.aspx)  
374 [ning\\_in\\_Mammography\\_Diagnostic\\_Accuracy.7.aspx](https://journals.lww.com/investigativeradiology/Fulltext/2017/07000/Deep_Learning_in_Mammography_Diagnostic_Accuracy.7.aspx)
- 375 14. Dembrower K, Wåhlin E, Liu Y, Salim M, Smith K, Lindholm P, et al. Effect of artificial  
376 intelligence-based triaging of breast cancer screening mammograms on cancer detection  
377 and radiologist workload: a retrospective simulation study. *The Lancet Digital Health*  
378 *[Internet]*. 2020 Sep 1 [cited 2021 Dec 3];2(9):e468–74. Available  
379 from: <http://www.thelancet.com/article/S2589750020301850/fulltext>
- 380 15. Buda M, Saha A, Walsh R, Ghate S, Li N, Święcicki A, et al. A Data Set and Deep Learning  
381 Algorithm for the Detection of Masses and Architectural Distortions in Digital Breast  
382 Tomosynthesis Images. *JAMA Network Open [Internet]*. 2021 Aug 2 [cited 2021 Dec  
383 3];4(8):e2119100–e2119100. Available  
384 from: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2783046>
- 385 16. Huang SH, Loh JK, Tsai JT, Houg MF, Shi HY. Predictive model for 5-year mortality after  
386 breast cancer surgery in Taiwan residents. *Chinese Journal of Cancer [Internet]*. 2017

- 387 Feb 27 [cited 2021 Dec 7];36(1):1–9. Available  
 388 from: <https://cancercommun.biomedcentral.com/articles/10.1186/s40880-017-0192-9>
- 389 17. Juwara L, Arora N, Gornitsky M, Saha-Chaudhuri P, Velly AM. Identifying predictive  
 390 factors for neuropathic pain after breast cancer surgery using machine learning.  
 391 International Journal of Medical Informatics. 2020 Sep 1;141:104170.
- 392 18. Lötsch J, Sipilä R, Dimova V, Kalso E. Machine-learned selection of psychological  
 393 questionnaire items relevant to the development of persistent pain after breast cancer  
 394 surgery. British journal of anaesthesia [Internet]. 2018 Nov 1 [cited 2021 Dec  
 395 3];121(5):1123–32. Available from: <https://pubmed.ncbi.nlm.nih.gov/30336857/>
- 396 19. Fu MR, Wang Y, Li C, Qiu Z, Axelrod D, Guth AA, et al. Machine learning for detection of  
 397 lymphedema among breast cancer survivors. mHealth [Internet]. 2018 May [cited 2021  
 398 Dec 3];4:17–17. Available from: <https://pubmed.ncbi.nlm.nih.gov/29963562/>
- 399 20. Lou SJ, Hou MF, Chang HT, Chiu CC, Lee HH, Yeh SCJ, et al. Machine Learning Algorithms  
 400 to Predict Recurrence within 10 Years after Breast Cancer Surgery: A Prospective Cohort  
 401 Study. Cancers [Internet]. 2020 Dec 1 [cited 2021 Dec 3];12(12):1–15. Available  
 402 from: <https://pubmed.ncbi.nlm.nih.gov/33348826/>
- 403 21. Myung Y, Jeon S, Heo C, Kim EK, Kang E, Shin HC, et al. Validating machine learning  
 404 approaches for prediction of donor related complication in microsurgical breast  
 405 reconstruction: a retrospective cohort study. Scientific Reports 2021 11:1 [Internet].  
 406 2021 Mar 10 [cited 2021 Dec 3];11(1):1–9. Available  
 407 from: <https://www.nature.com/articles/s41598-021-85155-z>
- 408 22. van Egdom LSE, Pusic A, Verhoef C, Hazelzet JA, Koppert LB. Machine learning with PROs  
 409 in breast cancer surgery; caution: Collecting PROs at baseline is crucial. The Breast  
 410 Journal [Internet]. 2020 Jun 1 [cited 2021 Dec 7];26(6):1213–5. Available  
 411 from: <https://onlinelibrary.wiley.com/doi/full/10.1111/tbj.13804>
- 412 23. Yang N, Zheng Z, Wang T. Model loss and distribution analysis of regression problems in  
 413 machine learning. PervasiveHealth: Pervasive Computing Technologies for Healthcare  
 414 [Internet]. 2019 [cited 2021 Dec 7];Part F148150:1–5. Available  
 415 from: <https://doi.org/10.1145/3318299.3318367>
- 416 24. Yap MH, Goyal M, Osman F, Martí R, Denton E, Juette A, et al. Breast ultrasound region  
 417 of interest detection and lesion localisation. Artificial Intelligence in Medicine. 2020 Jul  
 418 1;107:101880.
- 419 25. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable  
 420 machine learning can outperform Cox regression predictions and provide insights in  
 421 breast cancer survival. Scientific Reports 2021 11:1 [Internet]. 2021 Mar 26 [cited 2021  
 422 Dec 8];11(1):1–13. Available from: <https://www.nature.com/articles/s41598-021-86327-7>
- 423 26. Sidey-Gibbons C, Pfof A, Asaad M, Boukovalas S, Lin Y-L, Selber JC, et al. Development of  
 424 Machine Learning Algorithms for the Prediction of Financial Toxicity in Localized Breast  
 425 Cancer Following Surgical Treatment. JCO Clinical Cancer Informatics. 2021 Mar  
 426 25;(5):338–47.
- 427 27. Huber M, Kurz C, Leidl R. Predicting patient-reported outcomes following hip and knee  
 428 replacement surgery using supervised machine learning. BMC Medical Informatics  
 429 and Decision Making [Internet]. 2019 Jan 8 [cited 2021 Dec 7];19(1):1–13. Available  
 430 from: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-018-0731-6>
- 431 28. Chiew CJ, Liu N, Wong TH, Sim YE, Abdullah HR. Utilizing Machine Learning Methods for  
 432 Preoperative Prediction of Postsurgical Mortality and Intensive Care Unit Admission.  
 433 Annals of surgery [Internet]. 2020 Dec 1 [cited 2021 Dec 7];272(6):1133–9. Available  
 434 from: [https://journals.lww.com/annalsofsurgery/Fulltext/2020/12000/Utilizing\\_Machin  
 435 e\\_Learning\\_Methods\\_for.39.aspx](https://journals.lww.com/annalsofsurgery/Fulltext/2020/12000/Utilizing_Machine_Learning_Methods_for.39.aspx)  
 436  
 437

- 438 29. Lu S, Yan M, Li C, Yan C, Zhu Z, Lu W. Machine-learning-assisted prediction of surgical  
439 outcomes in patients undergoing gastrectomy. Chinese Journal of Cancer Research  
440 [Internet]. 2019 [cited 2021 Dec 7];31(5):797. Available from:  
441 /pmc/articles/PMC6856706/
- 442 30. Hale AT, Stonko DP, Brown A, Lim J, Voce DJ, Gannon SR, et al. Machine-learning analysis  
443 outperforms conventional statistical models and CT classification systems in predicting  
444 6-month outcomes in pediatric patients sustaining traumatic brain injury. Neurosurgical  
445 Focus [Internet]. 2018 Nov 1 [cited 2021 Dec 7];45(5):E2. Available  
446 from: <https://thejns.org/focus/view/journals/neurosurg-focus/45/5/article-pE2.xml>
- 447 31. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable  
448 machine learning can outperform Cox regression predictions and provide insights in  
449 breast cancer survival. Scientific Reports 2021 11:1 [Internet]. 2021 Mar 26 [cited 2021  
450 Dec 7];11(1):1–13. Available from: <https://www.nature.com/articles/s41598-021-86327-7>
- 451 32. Krogh A. What are artificial neural networks? Nature Biotechnology 2008 26:2 [Internet].  
452 2008 Feb [cited 2021 Dec 7];26(2):195–7. Available  
453 from: <https://www.nature.com/articles/nbt1386>
- 454 33. Zafeiris D, Rutella S, Ball GR. An Artificial Neural Network Integrated Pipeline for  
455 Biomarker Discovery Using Alzheimer’s Disease as a Case Study. Computational and  
456 Structural Biotechnology Journal [Internet]. 2018 [cited 2021 Dec 7];16:77. Available  
457 from: /pmc/articles/PMC6026215/
- 458 34. López-Martínez F, Núñez-Valdez ER, Crespo RG, García-Díaz V. An artificial neural  
459 network approach for predicting hypertension using NHANES data. Scientific Reports  
460 [Internet]. 2020 Dec 1 [cited 2021 Dec 7];10(1):10620. Available from:  
461 /pmc/articles/PMC7327031/
- 462 35. Ayer T, Alagoz O, Chhatwal J, Shavlik JW, Kahn CE, Burnside ES. Breast Cancer Risk  
463 Estimation with Artificial Neural Networks Revisited: Discrimination and Calibration.  
464 Cancer [Internet]. 2010 Jul 15 [cited 2021 Dec 7];116(14):3310. Available from:  
465 /pmc/articles/PMC2920215/
- 466 36. Loftus T, Filiberto A, Li Y, Balch J, Cook A, Tighe P et al. Decision analysis and  
467 reinforcement learning in surgical decision-making. Surgery [Internet]. 2020 [cited 31  
468 December 2021];168(2):253-266. Available from:  
469 [https://www.surgjournal.com/article/S0039-6060\(20\)30261-0/fulltext](https://www.surgjournal.com/article/S0039-6060(20)30261-0/fulltext)
- 470 37. Kanevsky J, Corban J, Gaster R, Kanevsky A, Lin S, Gilardino M. Big Data and Machine  
471 Learning in Plastic Surgery: A New Frontier in Surgical Innovation. Plastic and  
472 Reconstructive Surgery [Internet]. 2016 May 1 [cited 2021 Dec 3];137(5):890e–7e.  
473 Available  
474 from: [https://journals.lww.com/plasreconsurg/Fulltext/2016/05000/Big\\_Data\\_and\\_Mac  
475 hine\\_Learning\\_in\\_Plastic\\_Surgery\\_.45.aspx](https://journals.lww.com/plasreconsurg/Fulltext/2016/05000/Big_Data_and_Machine_Learning_in_Plastic_Surgery_.45.aspx)
- 476 38. Bouaud J, Pelayo S, Lamy JB, Prebet C, Ngo C, Teixeira L, et al. Implementation of an  
477 ontological reasoning to support the guideline-based management of primary breast  
478 cancer patients in the DESIREE project. Artificial Intelligence in Medicine. 2020 Aug  
479 1;108:101922.
- 480 39. Gu D, Su K, Zhao H. A case-based ensemble learning system for explainable breast  
481 cancer recurrence prediction. Artificial Intelligence in Medicine. 2020 Jul 1;107:101858.
- 482 40. Borsting E, Desimone R, Ascha M, Ascha M. Applied Deep Learning in Plastic Surgery:  
483 Classifying Rhinoplasty with a Mobile App. Journal of Craniofacial Surgery [Internet].  
484 2020 Jan 1 [cited 2021 Dec 3];31(1):102–6. Available  
485 from: [https://journals.lww.com/jcraniofacialsurgery/Fulltext/2020/01000/Applied\\_Deep  
486 Learning\\_in\\_Plastic\\_Surgery\\_.27.aspx](https://journals.lww.com/jcraniofacialsurgery/Fulltext/2020/01000/Applied_Deep_Learning_in_Plastic_Surgery_.27.aspx)
- 487

- 488 41. Murphy DC, Saleh DB. Artificial Intelligence in plastic surgery: What is it? Where are we  
489 now? What is on the horizon? Annals of The Royal College of Surgeons of England  
490 [Internet]. 2020 Aug 11 [cited 2021 Dec 3];102(8):577. Available from:  
491 /pmc/articles/PMC7538735/  
492 42. Zhu VZ, Tuggle CT, Au AF. Promise and limitations of big data research in plastic surgery.  
493 Annals of Plastic Surgery [Internet]. 2016 [cited 2021 Dec 3];76(4):453–8. Available  
494 from: [https://journals.lww.com/annalsplasticsurgery/Fulltext/2016/04000/Promise\\_and](https://journals.lww.com/annalsplasticsurgery/Fulltext/2016/04000/Promise_and_Limitations_of_Big_Data_Research_in.20.aspx)  
495 [Limitations of Big Data Research in.20.aspx](https://journals.lww.com/annalsplasticsurgery/Fulltext/2016/04000/Promise_and_Limitations_of_Big_Data_Research_in.20.aspx)  
496 43. Manlhiot C, van den Eynde J, Kutty S, Ross H. A Primer on the Present State and Future  
497 Prospects for Machine Learning and Artificial Intelligence Applications in Cardiology.  
498 Canadian Journal of Cardiology [Internet]. 2021 [cited 28 January 2022];. Available from:  
499 <https://www.sciencedirect.com/science/article/pii/S0828282X2100903X>  
500 44. Roscher R, Bohn B, Duarte MF, Garcke J. Explainable Machine Learning for Scientific  
501 Insights and Discoveries. IEEE Access. 2020;8:42200–16.  
502 45. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable  
503 machine-learning predictions for the prevention of hypoxaemia during surgery. Nature  
504 Biomedical Engineering 2018 2:10 [Internet]. 2018 Oct 10 [cited 2021 Dec 8];2(10):749–  
505 60. Available from: <https://www.nature.com/articles/s41551-018-0304-0>  
506 46. Sheller M, Edwards B, Reina G, Martin J, Pati S, Kotrotsou A et al. Federated learning in  
507 medicine: facilitating multi-institutional collaborations without sharing patient data.  
508 Scientific Reports [Internet]. 2020 [cited 28 January 2022];10(1). Available from:  
509 <https://www.nature.com/articles/s41598-020-69250-1#citeas>  
510 47. O’Neill AC, Yang D, Roy M, Sebastiampillai S, Hofer SOP, Xu W. Development and  
511 Evaluation of a Machine Learning Prediction Model for Flap Failure in Microvascular  
512 Breast Reconstruction. Annals of Surgical Oncology 2020 27:9 [Internet]. 2020 Mar 9  
513 [cited 2021 Dec 8];27(9):3466–75. Available  
514 from: <https://link.springer.com/article/10.1245/s10434-020-08307-x>  
515  
516

517

518

519

520

521

522

523

524

525 **Figure legend:**

526 **Figure 1.** PRISMA Flow Chart

527 **Figure 2.** Risk of bias diagram

528 **Figure 3:** Challenges and recommendation in Machine Learning research within Breast Surgery.

529 **Table legends:**

530 **Table 1.** studies included assessing the use of machine learning in breast surgery

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545 **Table 1: studies included assessing the use of machine learning in breast surgery**

Study	Year	Study design	Type of Procedure	Category of machine learning method utilised	Population Number	Method of machine learning implemented	Main reported outcomes
Becker et al	2017	NM, NR, NP	Mammography	Image analysis	3228	<ul style="list-style-type: none"> <li>Patients experiencing mammography at the institution were chosen to form the dataset - this was used to train the neural network.</li> <li>An external dataset from the Breast Cancer Digital Repository was used for testing. This was appraised by three radiologists.</li> </ul>	<ul style="list-style-type: none"> <li>One radiologist showed nearly equivalent performance to the network (0.83, <math>p = 0.17</math>) and the other two performed significantly better (0.91 and 0.94 respectively, <math>p &lt; 0.016</math>).</li> <li>The neural network's performance of 0.82 did not differ significantly between radiologist performance. The neural network behaved less specifically and more sensitively compared to humans throughout.</li> </ul>
Dembrower et al	2020	M, NR, NP	Mammography	Screening	7364	<ul style="list-style-type: none"> <li>A cohort of mammograms were made selected in a way that mimicked frequency in reality. The AI cancer detector algorithm used had been pre-trained.</li> <li>The software generates underlying image-level prediction scores for tumour presence. From this, cutoff points were established to insert patients into two novel work streams - on missed and additionally detected cancer.</li> </ul>	<ul style="list-style-type: none"> <li>For 60%, 70% or 80% of women possessing the lowest AI scores in the negative radiologist stream, 0.3% (95% CI 0.0–4.3) / 2.6% (1.1–5.4) of screen-detected cancers would be missed respectively. For the 1% or 5% of women possessing highest AI scores in the 'enhanced assessment' stream, 12% or 27% of subsequent interval cancers, respectively, and 14% or 35% of next-round screen-detected cancers, respectively, may have also been able to be detected additionally.</li> </ul>
Buda et al	2021	M, NR, NP	Digital breast tomosynthesis	Deep learning algorithm	16802	<ul style="list-style-type: none"> <li>16802 digital breast tomosynthesis examinations that had 1 reconstruction view between August 26, 2014 and January 29, 2018 were analysed.</li> <li>These were subdivided into 4 groups and further split into training and testing sets for the evaluation of the deep learning model.</li> </ul>	<ul style="list-style-type: none"> <li>The deep learning model trained reached a breast-based sensitivity at 65% (39 of 60; 95% CI, 56%-74%) on the test set. This was at 2 false positives per DBT volume.</li> </ul>
Huang et al	2017	M, NR, NP	Breast cancer surgery	Predictive model	3632	<ul style="list-style-type: none"> <li>This study compared three models (MLR, Cox &amp; ANN) for 3632 post-operative breast cancer patients between 1996 and 2010.</li> <li>An estimation dataset trained the model, and a validation dataset helped evaluate the performance of the model. Sensitivity analysis allowed the comparison of input variables for the model's predictions.</li> </ul>	<ul style="list-style-type: none"> <li>The ANN model overall performed best over the MLR and Cox models for predicting 5-year breast cancer mortality post-operatively.</li> <li>Age, Charlson comorbidity index (CCI), chemotherapy, radiotherapy, hormone therapy, breast cancer surgery volumes of hospital and breast cancer surgery volumes of surgeon were significant associations with 5 year breast cancer surgery, the latter was the most significant.</li> </ul>
Juwara et al	2020	NM, NR, P	Breast cancer surgery	Predictive model	195	<ul style="list-style-type: none"> <li>6 machine learning algorithms (least square, ridge, elastic net, random forest, gradient boosting and neural net) aided primary analysis of the identification of predictors of DN4-interview score (index for neuropathic pain).</li> <li>Models were compared. A logistic classification model was created for neuropathic pain using the predictor outcomes of the primary analysis.</li> </ul>	<ul style="list-style-type: none"> <li>Anxiety, type of surgery, preoperative baseline pain and acute pain on movement predicted DN4-score most pertinently. Anxiety had the most significant association with neuropathic pain.</li> <li>The least square regression model compared well with the random forest model and neural network model. The Gradient boosting model performed better than all the other models.</li> </ul>

<b>Lotsch et al</b>	2018	NM, NR, P	Breast cancer surgery	Supervised machine learning prediction	1000	<ul style="list-style-type: none"> <li>ML helped establish a short questionnaire to identify pain to the same predictive level as pre-existing full-form questionnaires.</li> <li>The predictors were trained via the full set of items in Beck's Depression Inventory (BDI), Spielberger's State-Trait Anxiety Inventory (STAI), and the State-Trait Anger Expression Inventory (STAXI-2).</li> <li>ML extracted features of these to create predictors with a lower item number.</li> </ul>	<ul style="list-style-type: none"> <li>A 7-item set produced via ML that comprised 10% of the original questions from the STAI and BDI respectively performed the same as the full questionnaires in predicting development of persistent postsurgical pain.</li> </ul>
<b>Fu et al</b>	2018	M, NR, NP	Breast cancer surgery	Detection	355	<ul style="list-style-type: none"> <li>A mobile health system allowed data to be collected based on real-time symptom reporting.</li> <li>Both statistical and ML processes were employed for data analysis. Regarding the latter, 5 classical algorithms were compared: Decision Tree of C4.5, Decision Tree of C5.0, gradient boosting model (GBM), artificial neural network (ANN), and support vector machine (SVM).</li> </ul>	<ul style="list-style-type: none"> <li>Using ML to compare different algorithms is a viable concept. Artificial neural network (ANN) performed best for detecting post-operative development of lymphedema (accuracy - 93.75%, sensitivity - 95.65% and specificity - 91.03%).</li> </ul>
<b>Lou et al</b>	2020	M, NR, NP	Breast conservation surgery, modified reconstructive mastectomy, mastectomy with reconstruction	Predictive model	1140	<ul style="list-style-type: none"> <li>The cases in this study were divided into a training dataset to develop the ML model, a testing dataset for internal validation and an externally validating dataset.</li> <li>After training, outputs of the model were taken for each training set. Accuracy in predicting breast cancer recurrence within 10 years was compared.</li> </ul>	<ul style="list-style-type: none"> <li>The ANN model performed significantly best of all models based on sensitivity, specificity, PPV, NPV, accuracy, and AUROC values.</li> <li>Surgeon volume followed by hospital volume and tumour grade were, in that order, best predictors of recurrence of breast cancer within 10 years.</li> </ul>
<b>Myung et al</b>	2021	NM, NR, NP	Microsurgical unilateral breast reconstruction: muscle-sparing type TRAM and DIEP (muscle sparing type 3) abdominal flaps	Predictive model	568	<ul style="list-style-type: none"> <li>Neuralnet, and RSNNs machine learning packages were applied to compare prediction accuracy, sensitivity, specificity, and predictive power (AUC) for predicting factors that raise abdominal flap donor site complications (against logistic regression).</li> <li>13 variables suggested to influence donor site complication rates were evaluated.</li> </ul>	<ul style="list-style-type: none"> <li>Neuralnet performed most optimally of all the packages.</li> <li>Fascial defect, history of diabetes, muscle sparing type, and presence or absence of adjuvant chemotherapy all significantly affected complication rate of donor sites.</li> <li>High-risk group complication rates were significant compared to the low-risk group upon statistical analysis.</li> </ul>
<b>Van Egdome et al</b>	2020	NM, NR, NP	Breast cancer surgery	Predictive model	764	<ul style="list-style-type: none"> <li>Various patient data variables were available.</li> <li>ML methods (general linear model regression (GLM), support vector machines (SVM), single-layer artificial neural networks (ANN), and deep learning (DL)) evaluated presurgical prognostic factors of age, medical status, tumour characteristics, and (neo)adjuvant treatment indications or treatment characteristics.</li> </ul>	<ul style="list-style-type: none"> <li>No relationship was determined between predictors and outcomes, rendering the model akin to the outcomes' respective population prevalence. Combining variables and, simultaneously, reducing dimensions, did not yield significant changes.</li> </ul>
<b>Yang et al</b>	2021	NM, NR, NP	Breast cancer surgery	Predictive model	1061	<ul style="list-style-type: none"> <li>This study posed a predictive model of breast cancer recurrence based on clinical, nominal and numeric features.</li> <li>6 features from an initial dataset were identified for further processing and resampling.</li> <li>AdaBoost and cost-sensitive learning packages predicted the</li> </ul>	<ul style="list-style-type: none"> <li>AdaBoost reaches an accuracy of 0.973 and sensitivity of 0.675. A combination of AdaBoost and cost-sensitive learning poses a model with a reasonable accuracy of 0.468 and very high sensitivity of 0.947. Hence, the model is can be used to support early intervention.</li> </ul>

						risk of recurrence and evaluated performance.	
<b>Yap et al</b>	2020	M, NR, NP	Breast ultrasound detection	Image analysis	469	<ul style="list-style-type: none"> <li>The study employs a deep learning method for breast ultrasound ROI detection and lesion localisation. Transfer learning is used due to unavailability of datasets and a novel 3-channel artificial RGB method is applied for performance improvement.</li> <li>This proposed method is evaluated and compared using an individual and composite dataset.</li> </ul>	<ul style="list-style-type: none"> <li>Faster-RCNN outperformed a computer vision object detection algorithm indicating viability for use in BUS lesion localisation.</li> <li>IoU (equivalent to Dice Coefficient Index) should be used in lesion detection due to its reliability.</li> </ul>
<b>Moncada-Torres et al</b>	2021	M, NR, P	Breast-conserving surgery, mastectomy	Predictive model	36658	<ul style="list-style-type: none"> <li>Data of patients who underwent curative breast surgery was used to compare the performance of Cox proportional hazards analysis (CPH) with ML models (Random Survival Forests, Survival Support Vector Machines and Extreme Gradient Boosting [XGB]) for survival predictions.</li> </ul>	<ul style="list-style-type: none"> <li>ML models perform to at least the same standard as classical CPH regression and even better for some models (XGB). Furthermore, Shapley Additive Explanation (SHAP) values were used successfully as a form of explainable machine learning to provide detail on how the models' predictions are made.</li> </ul>
<b>Sidey-Gibbons et al</b>	2021	NM, NR, P	Breast cancer surgery	Predictive model	611	<ul style="list-style-type: none"> <li>A set of oML algorithms (neural network, regularized linear model, support vector machines, and a classification tree) were trained and tested for making predictions of financial toxicity in a dataset.</li> <li>The data were split into samples for the training and testing sets, prior to assessment of predictive performance.</li> </ul>	<ul style="list-style-type: none"> <li>ML packages accurately predicted financial toxicity in this context demonstrating an AUROC of 0.85, accuracy of 0.82, sensitivity of 0.85, and specificity of 0.81.</li> <li>Neoadjuvant therapy and autologous reconstruction were ascertained as key indicators of financial toxicity.</li> <li>Radiation and tumour grade showed no effect.</li> </ul>

546 *M: multicentre; NM: non-multicentre; R: randomized; NR: non-randomized; P: prospective, NP: non-prospective*

547

548

549

550

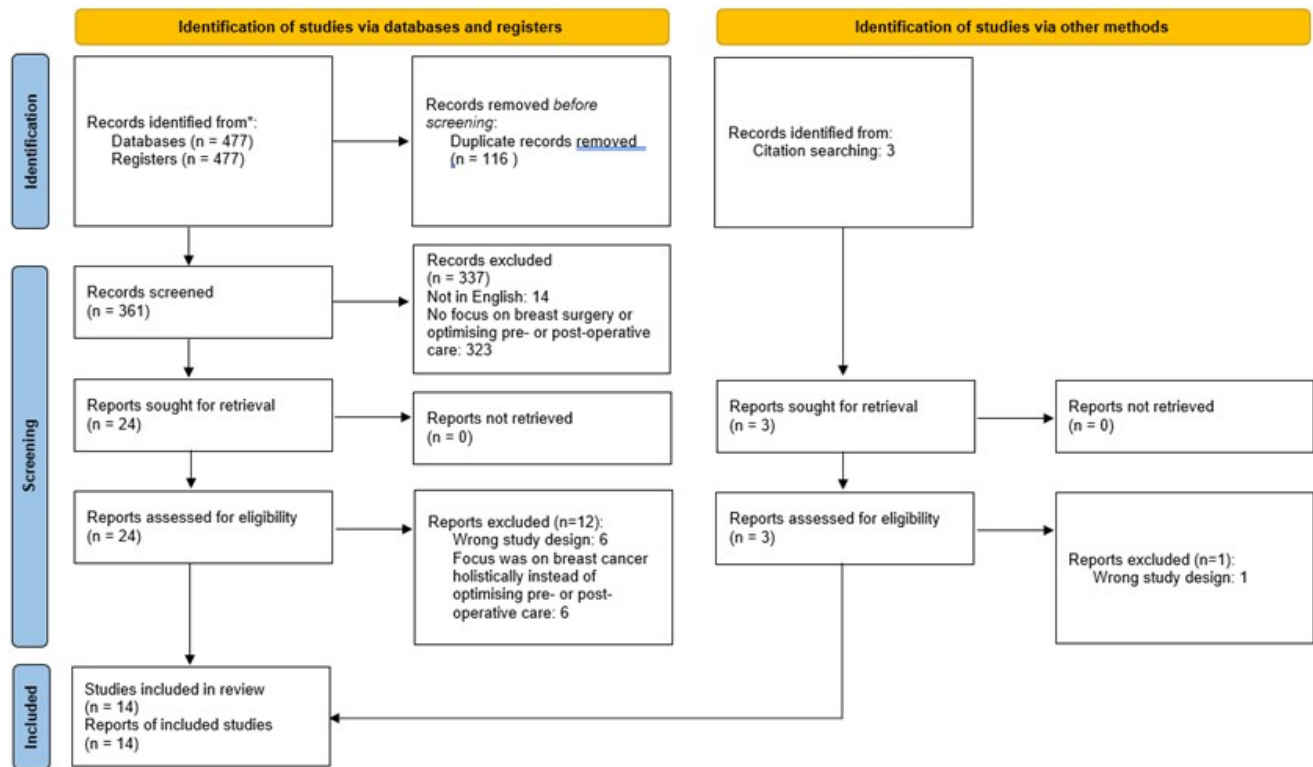
551

552

553

554

555



556

557

558

559

560

561

Study	Risk of bias domains					Overall
	D1	D2	D3	D4	D5	
Becker et al	+	+	+	+	-	-
Dembrower et al	+	+	+	+	+	+
Buda et al	-	+	+	+	+	-
Huang et al	+	+	+	+	+	+
Juwara et al	+	+	+	+	-	-
Lotsch et al	+	+	+	+	-	-
Fu et al	-	+	+	+	+	-
Lou et al	+	+	+	+	+	+
Myung et al.	+	+	+	+	+	+
Van Egdom et al	-	+	-	-	-	X
Xu et al	-	+	+	-	-	X
Yang et al	+	+	+	+	+	+
Yap et al	-	+	+	+	+	-
Gu et al	X	-	+	-	+	X
Bouaud et al	+	+	+	+	-	-
Moncada-Torres et al	+	+	+	+	-	-

Domains:  
 D1: Bias arising from the randomization process.  
 D2: Bias due to deviations from intended intervention.  
 D3: Bias due to missing outcome data.  
 D4: Bias in measurement of the outcome.  
 D5: Bias in selection of the reported result.

Judgement  
 High  
 Some concerns  
 Low

# Recommendations for the Adoption of Machine Learning in Breast Surgery

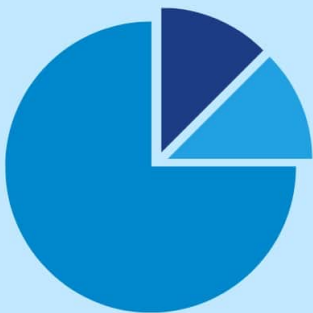


## Challenge: Obtaining large-scale evidence

- Currently many case studies and proof of concept descriptions
- To overcome sample bias and improve representation for the best evidence
- Use **cohort studies** and large **randomised control trials**

## Challenge: Inviting stakeholders

- Recognise the value of involving all stakeholders in design of ML processes
- Use **focus groups** and **expert interviews** when designing study protocols
- Improve synergy between ML and surgeons and promote understanding and leverage expertise



## Challenge: Utilising the best dataset

- ML should mirror the target population as best as possible.
- Utilise **readily available data** or **combining existing datasets** in resource-poor settings.
- Use **transformation techniques** such as **oversampling** to reduce data imbalances

566

567

568

569

570

571

572

573

574

575

576