How Do Artificial Intelligence Applications in Breast Cancer Impact Health Equity, Particularly in Diagnosis, Treatment, and Outcomes of Diverse Groups: A Critical Literature Review

Kethakee Madhavika Andaraweera, Ratendra Chauhan, Arpana Verma

Submitted to: Journal of Medical Internet Research
on: October 19, 2025

Table of Contents

How Do Artificial Intelligence Applications in Breast Cancer Impact Health Equity, Particularly in Diagnosis, Treatment, and Outcomes of Diverse Groups: A Critical Literature Review

Kethakee Madhavika Andaraweera[1*] MBBS, MPH; Ratendra Chauhan[2*] BPT, MPT, MPH; Arpana Verma[1] MBChB, MPH, PhD, FFPH

[1]Division of Population Health, Health Services Research & Primary Care Faculty of Biology, Medicine and Health University of Manchester Manchester GB
[2] University of Manchester Manchester GB
[*]these authors contributed equally

Corresponding Author:
Kethakee Madhavika Andaraweera MBBS, MPH
Division of Population Health, Health Services Research & Primary Care
Faculty of Biology, Medicine and Health
University of Manchester
Oxford Road
Manchester
GB

Abstract

Background: The integration of Artificial Intelligence (AI) in breast cancer care has the potential to impact health equity, yet the extent of this impact remains underexplored. Understanding how AI applications influence health disparities, including racial, ethnic, and socio-economic factors, is crucial for developing effective and equitable interventions.

Objective: This review aims to provide a comprehensive analysis of current evidence on the impact of AI technologies on health equity in breast cancer (BC) care. It focuses on how AI influences disparities in diagnosis, treatment, and outcomes among different racial, ethnic, and socio-economic groups, and critically assesses the quality and robustness of studies evaluating AI interventions intended to address these inequities.

Methods: Adhering to PRISMA 2020 guidelines for Systematic Reviews, this Critical Literature Review analyzed studies published between January 2019 and August 2024. A comprehensive literature search across Medline, Embase, Web of Science, and Scopus was conducted to identify peer-reviewed research addressing AI applications in breast cancer and their impact on health equity. Studies were evaluated for their focus on health equity outcomes, and methodological rigor. The quality of each study was assessed using the Public Health Critical Appraisal Checklist (PHCA) and the Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML) checklist.

Results: The review included seven studies. The findings of the study highlighted three main themes of AI's role in reducing healthcare inequalities: forecasting biological and socio-economic factors affecting outcomes, enhancing decision-making fairness, and use of AI to reduce healthcare inequalities in a global context.

Conclusions: The studies highlight AI's potential in addressing health inequalities in breast cancer through its capacity to analyze complex datasets, incorporate race/ethnicity-specific models, and function in resource-limited settings. However, its impact depends heavily on data quality, dataset representativeness, and the ability to mitigate inherent model biases. Therefore, policymakers should focus on improving dataset inclusivity, establishing benchmark datasets, standardizing data collection, and addressing data poverty in low- and middle-income countries. Future researchers should ensure dataset representativeness and generalizability, conduct prospective or randomized controlled trials to assess real-world applications, and carefully select predictor variables through multidisciplinary collaboration. Adherence to standard reporting guidelines is also essential to enhance reproducibility and ensure the reliability of AI applications in breast cancer.

(JMIR Preprints 19/10/2025:86132)
DOI: https://doi.org/10.2196/preprints.86132

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

How Do Artificial Intelligence Applications in Breast Cancer Impact Health Equity, Particularly in Diagnosis, Treatment, and Outcomes of Diverse Groups: A Critical Literature Review

Abstract

Background: The integration of Artificial Intelligence (AI) in breast cancer care has the potential to

impact health equity, yet the extent of this impact remains underexplored. Understanding how AI applications influence health disparities, including racial, ethnic, and socio-economic factors, is crucial for developing effective and equitable interventions.

Objectives: This review aims to provide a comprehensive analysis of current evidence on the impact of AI technologies on health equity in breast cancer (BC) care. It focuses on how AI influences disparities in diagnosis, treatment, and outcomes among different racial, ethnic, and socio-economic groups, and critically assesses the quality and robustness of studies evaluating AI interventions intended to address these inequities.

Methodology: Adhering to PRISMA 2020 guidelines for Systematic Reviews, this Critical Literature Review analyzed studies published between January 2019 and August 2024. A comprehensive literature search across Medline, Embase, Web of Science, and Scopus was conducted to identify peer-reviewed research addressing AI applications in breast cancer and their impact on health equity. Studies were evaluated for their focus on health equity outcomes, and methodological rigor. The quality of each study was assessed using the Public Health Critical Appraisal Checklist (PHCA) and the Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML) checklist.

Results: The review included seven studies. The findings of the study highlighted three main themes of AI's role in reducing healthcare inequalities: forecasting biological and socio-economic factors affecting outcomes, enhancing decision-making fairness, and use of AI to reduce healthcare inequalities in a global context.

Conclusion: The studies highlight AI’s potential in addressing health inequalities in breast cancer through its capacity to analyze complex datasets, incorporate race/ethnicity-specific models, and function in resource-limited settings. However, its impact depends heavily on data quality, dataset representativeness, and the ability to mitigate inherent model biases. Therefore, policymakers should focus on improving dataset inclusivity, establishing benchmark datasets, standardizing data

collection, and addressing data poverty in low- and middle-income countries. Future researchers should ensure dataset representativeness and generalizability, conduct prospective or randomized controlled trials to assess real-world applications, and carefully select predictor variables through multidisciplinary collaboration. Adherence to standard reporting guidelines is also essential to enhance reproducibility and ensure the reliability of AI applications in breast cancer.

1. Introduction

1.1 Background

1.1.1Breast Cancer and Disparities

Breast cancer (BC) refers to a malignant tumor that originates in the cells of the different parts of the breast[1]. It is the most common cancer and the leading cause of cancer death among women globally, accounting for approximately 600,000 in 2022[2].

While advances in early detection and treatment have improved outcomes, especially in high income countries (HICs)[2,3], BC care is marked with significant disparities across racial, ethnic, and socio-economic groups. Numerous studies[4–9] have documented these inequalities, highlighting differences in BC incidence, screening, diagnosis, treatment, and outcomes. Among these studies, the most frequent themes were racial disparities within countries and geographical disparities across the globe. For instance, it is widely stated that even though European descent white women have higher incidences of BC, minority populations, particularly African American and Hispanic women, are more likely to be diagnosed at later stages, hence having poorer survival rates compared to their white counterparts[4,10]. Also, it is found that HIC have seen significant reductions in BC mortalities compared to low-middle-income countries (LMICs) in the last twenty years[11]. These disparities are grounded in wider determinants of health such as access to healthcare, geographic

location, socio-economic status, and systemic biases within the healthcare systems as well as biological determinants. Delays in diagnosis and treatment, compounded by unequal access to quality care, contribute to the lower survival rates observed among disadvantaged populations [6,8].

1.1.2 AI Integration into Breast Cancer Care

Artificial Intelligence (AI) I is broadly defined as the “technology that enables computers and machines to simulate human learning, comprehension, problem-solving, decision making, creativity and autonomy”[12]. AI has been progressively integrated into various applications of healthcare[13,14] with significantly increasing utilisation in cancer care, including BC. This

includes cancer research, expediting cancer screening, detection, and diagnosis via data and imaging analysis, treatment recommendations, and personalized care planning as well as drug discovery and improving cancer surveillance[15,16]. The use of AI in BC care has grown rapidly, with AI models showing the potential to enhance diagnostic accuracy, predict treatment outcomes, improve overall patient care, and facilitate personalized care, especially when utilizing machine learning (ML) and deep learning (DL)[17–19]. For example, a review found that in a mammography-based screening program using AI, the screening performance improved while reducing the workload[20] Additionally, another observational study has shown that AI algorithms outperformed the standard clinical risk model for predicting the five year risk of BC[21].

1.1.3 AI and Health Equity

Health equity refers to the absence of unjust, preventable, or correctable differences among people, irrespective of their social, economic, demographic, or geographic background, including factors like gender, ethnicity, disability, or sexual orientation. Achieving health equity involves addressing

social determinants of health, shaped by structural factors such as political, legal, and economic systems, as well as social norms and institutional processes, creating inequities in the conditions of people's lives, directly influencing health outcomes, and perpetuating disparities[22,23].

With the increasing integration of AI into healthcare and cancer care[24–27], the impact on health equity is highly discussed. While literature supports the fact that AI could be a powerful tool for reducing disparities in cancer care they also raise concerns of potential new disparities and exacerbation of existing disparities[28–30]. For example, Mema and McGinty[31] discuss how AI can help understand and address disparities in BC outcomes by highlighting its potential to function as a real-time auditor, identifying unconscious bias in patient encounters and related decision-making, enhancing access to cancer care for underserved populations, and lowering administrative and delivery costs.

Conversely, studies such as those by Viswanathan et al[32] and Paulus and Kent[33], caution that AI could exacerbate existing disparities. They argue that AI models trained on data from specific racial or gender groups may perform poorly for others, particularly impacting racial and ethnic minorities, low-income individuals, and those in geographically underserved areas. Also, they stated that algorithms developed with data from predominantly white or resource rich populations may result in less accurate predictions and increased inequities in healthcare access and outcomes for these underrepresented groups[29,32–36].

1.2 Study Rationale

However, a major limitation across the studies from both sides of the argument is the presence of geographical bias. For example, the reviews by Istasy et al.[30] predominantly included articles published in the United States, while studies by Mema and McGinty[31] and Yu and Zhai[36] focused largely on studies and examples from North America, the United Kingdom (UK), and Europe. This creates a significant risk of generalization bias which limits the applicability of their

conclusions in a broader global context.

In addition, much of the existing literature in this area consists of theoretical work, commentary, and narrative literature[28,29,31–34] which provides comparatively low reliability according to the hierarchy of evidence[37]. Moreover, A 2021 study demonstrated that the deployment of AI in oncology remains in its pre and early stages with most studies still focused on retrospective study validation[38] rather than prospective or real-time clinical implementation.

Given these limitations, there is a clear need for a more globally inclusive and methodologically rigorous review that prioritizes studies with higher levels of evidence, captures novel research on real-world implementations, and emphasizes equity-oriented outcomes to better inform future research and policy implementations in the field.

1.3. Research Gap

Despite the growing use of AI in BC care[39] and well-documented disparities in the field as seen above, there is a notable gap in research specifically examining AI's impact on health equity in BC. To date, there is one study reviewing AI's broader role in cancer care equity[30], and none focuses on BC specifically. Addressing this gap is crucial to ensure that AI benefits are equitably distributed and do not exacerbate existing disparities in BC outcomes. This study aims to fill this important gap, contributing to the development of more inclusive and ethical AI applications in BC care.

1.4 Research Question, Aim, and Objectives.

1.4.1 Aim and Objectives.

The overarching aim of this review is to provide a comprehensive overview and critical analysis of the current evidence on the impact of AI applications on health equity in BC care.

The objectives of this review are as follows:

1. To explore how AI technologies influence disparities in BC diagnosis, treatment, and outcomes among different racial, ethnic, and socio-economic groups.
2. To critically assess the methodological quality and robustness of studies evaluating AI interventions aimed at mitigating disparities in BC care.

1.4.2 Research Question

The central research question guiding this review is: How do AI applications in BC care impact health equity, particularly concerning diagnosis, treatment, and outcomes among diverse racial, ethnic, and socio-economic groups?

2. Review Methods

2.1 Search Strategy

This Critical Literature Review (CLR) was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines[40].
To gather relevant results on the research question, a literature search was carried out on August 23, 2024, across four online databases: Medline, Embase (accessed via OVID), Web of Science, and Scopus. It is important to acknowledge grey literature was not examined due to the limitations of the CLR, which may have led to the exclusion of some relevant studies.

The search strategy was based on the PICOS framework as it allows for capturing the key elements of the research question[41]. The breakdown of the research question according to the PICOS framework is detailed in Table 1.

Table 1: PICOS framework of the research question.

Population of Interest (P):	Breast cancer patients. (No geographic limitations)
Intervention (I):	The application of AI algorithms.

Comparison (C)	None
Outcome (O)	The impact on health equity.
Study type (S)	Any

To operationalize the search, variants of keywords related to AI, Health Equity, and BC were explored. The keywords were identified by scoping the grey literature on similar topics and concepts before initializing the search.

To ensure thorough coverage, the initial search results were not limited by period or geographic location. The language criterion restricting the final eligibility screening to studies published in English was applied only during the screening stage and did not limit the initial search results.
The detailed search strategies and results for all the databases are provided in S1 Appendix.

2.2 Study Selection for the Literature Review

A total of 211 records were identified through the search strategy described above. After deduplication, the records were screened by two independent reviewers evaluating their titles and abstracts. Records were selected if they referenced all three core concepts: AI, health equity, and BC, or related terms. Those that did not meet this criterion were excluded. Reports selected during this phase were then retrieved, and any that could not be accessed were excluded from further analysis. The full texts of the successfully retrieved articles were subsequently assessed by the two reviewers for eligibility using predefined inclusion and exclusion criteria. All conflicts regarding the selection of articles at each and every step were resolved through independent review by a third reviewer and discussion among the three reviewers.

2.2.1 Eligibility Criteria

Given the rapid advancements in the use of AI in oncology since 2019[42], particularly following

the COVID-19 pandemic[43,44], articles published before 2019 were considered potentially outdated and were therefore excluded.

Although a small percentage (0.5–1% %) of male patients are diagnosed with BC annually[2], and transgender individuals may also be affected[45], the review aimed to minimize population variability. Therefore, studies that did not specifically focus on female individuals were excluded. Age restrictions were not applied due to the limited number of women diagnosed before the age of forty-five[46].

Articles were excluded if they addressed cancer generally rather than specifically focusing on BC. Additionally, studies were excluded if their outcome measures centered on secondary consequences of BC, such as side effects related to treatment, to avoid introducing variability. Conversely, studies with primary outcomes related to BC diagnosis, treatment, or follow-up were included, regardless of discipline, such as epidemiology, pathology, radiology, surgery, or oncology, given the interrelated nature of these fields in BC care.

Letters, opinions, perspectives, editorials, and abstracts from poster presentations or conferences were excluded due to their lower position in the hierarchy of evidence[37].

In the context of this study, the inclusion criteria for AI required that AI be used algorithmically as an intervention, therefore, studies focusing on other aspects of digital health, such as telehealth or big data analytics, or where AI was not the primary intervention, were excluded.

Articles were excluded if they did not include health equity as a primary or secondary outcome, did not address health inequity in their study rationale, or lacked adequate references to health equity.

The inclusion and exclusion criteria are detailed in Table 2.

Table 2: Detailed inclusion and exclusion criteria.

	Inclusion Criteria	Exclusion Criteria
Population	Female individuals.	Biologically male individuals, transgender individuals.
Intervention	Use of AI algorithms as the primary intervention.	Insufficient relevance to AI. Digital health interventions not involving AI (e.g., chatbots, telemedicine, big data analysis); AI as a secondary intervention.
Breast Cancer	Studies related to breast cancer diagnosis, treatment, or follow-up across disciplines (e.g., epidemiology, pathology, radiology, surgery, oncology).	Insufficient relevance to breast cancer**.** Studies focusing on secondary outcomes related to breast cancer or its treatments (side effects). Studies focusing on multiple types of cancer including breast cancer.
Health Equity	Explicitly addressed as a primary or secondary outcome, or within the study rationale.	Insufficient relevance to health equity. Lack of reference to health equity or insufficient discussion of equity.
Study type	Randomized controlled trials, cohort studies, case control studies, diagnostic studies, cross-sectional studies, systematic reviews.	Opinion pieces, editorials, perspective commentaries, poster abstracts, conference abstracts, narrative reviews.

2.3 Quality Assessment

Due to the inclusion of studies from both public health and health informatics and the heterogeneity of the study types, the quality of each study was assessed using the Public Health Critical Appraisal Checklist (PHCA)[47] and the Ruling Out Bias Using Standard Tools in Machine

Learning (ROBUST-ML) checklist[48]. The PHCA checklist was used to analyze the study characteristics and implications of the studies, while the ROBUST-ML checklist assessed the rigor of the AI models employed. Both checklists are provided in the Supporting Information section.

2.4 Data Extraction and Analysis

The data extraction and analysis process was conducted by a single reviewer and peer-reviewed by another. The process comprised both descriptive and qualitative components. Descriptively, data was extracted on the institution of the senior author, peer reviewed status, study type, sample size, database and data collection period, and type of AI utilized. Qualitatively, the articles were analyzed for health equity aspects, including the study's rationale, objectives, and interpretation of findings related to health equity. Findings related to equity were then used to group the data for thematic analysis.

3. Results

3.1 Study Selection

A total of seven records were ultimately included in the review. The process by which these records were selected is detailed in Figure 2.

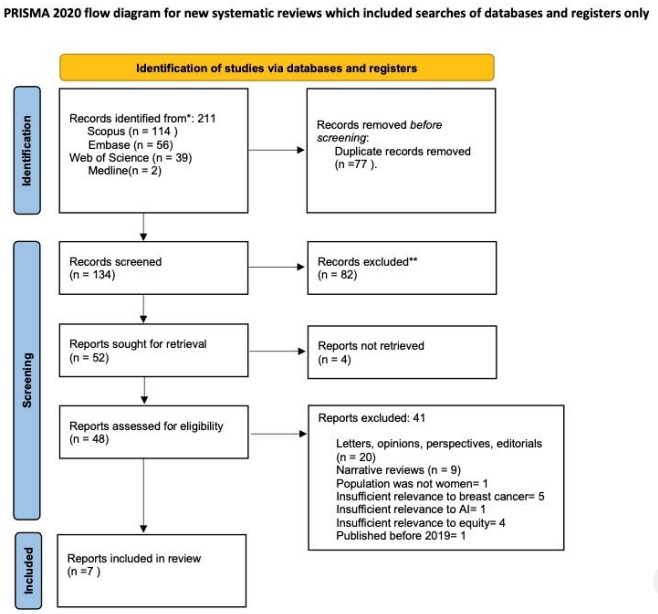

Figure 2: Selection flow of the studies.

3.2 Summary of Studies and Study Characteristics.

All the reviewed studies were conducted in the United States (U.S.), with research affiliated exclusively with major academic health centers. Notably, four studies[49–52] involved collaborations with non-governmental or governmental institutions. This concentration introduces potential biases related to resource access, geographic focus, and academic or policy-driven priorities, which may limit the generalizability of the findings to more diverse or global populations, especially with only one study[53] utilizing a database outside of the United States.

Dong W. et al.[50], Egleston et al.[51], and Park et al.[54] all utilize the Surveillance, Epidemiology and End Results (SEER) registries[55,56], which are widely recognized as the gold standard for cancer surveillance in the United States[57], thus enhancing the comprehensiveness and generalization ability of their data. Stone et al.[49] equip The Cancer Genome Atlas Program (TCGA) database[58], another robust public resource offering comprehensive cancer research data[59]. In contrast, Dell'Aquila et al.[60], Gao W. et al.[53], and Soltan and Washington[52] rely on data from single institutions or smaller, institution-specific databases[61,62], which may limit the generalizability of their findings due to less representativeness and less comprehensive demographic

information.

Four studies[50,51,53,54] leverage large sample sizes exceeding 200,000, significantly enhancing the statistical power and reliability of their findings. Soltan and Washington[52] also utilize a substantial dataset with over 10,000 images. In contrast, Dell'Aquila et al. [60] has a much smaller sample size, which may limit the strength of the subgroup analyses and AI model accuracy. Stone et al.[49], with the smallest sample size of 200, face challenges in statistical power, risking overfitting and biased results, particularly for deep learning models like ResNet the study utilize, that require larger samples for increased generalizability[48].

Although all studies were published after 2020, they analyzed data collected prior to 2020, with Egleston et al. using data dating back as far as 1992[51]. This introduces temporal bias[48] due to potential changes in medical treatments, diagnostic technologies, and healthcare policies over time. Additionally, Stone et al. do not specify the dates of sample collection[49], raising concerns that their images may not reflect current clinical practices, further limiting study generalizability. A detailed summary of the study characteristics is depicted in table 3.

Table 3: Summary of the study characteristics.

Study	**Institution by affiliation of first author.**	**Study Type**	**Sample size**	**Database and Data Collection Period.**	**AI Techniques Used (Algorithm/Model)**
Dell'Aquila et al. 2024 [60]	Albert Einstein College of Medicine, New York, U.S.	Retrospective cohort study with ML model development.	n=475	Institutional cancer registry and EMR database (2012-2021)	1.Machine Learning – Logistic Regression 2.Neural Network (NN), 3.Random Forest (RF), 4.Gradient Boosted Regressor (GBR)

Dong W. et al. 2022[50]	Case western Reserve University School of Medicine, Ohio, U.S.	Cross-sectional study with ML analysis.	n= 812,048	SEER database (2009-2018)	Machine Learning – 1.Classification and Regression Tree (CART), 2.Random Forest (RF)
Egleston et al. 2023[51]	Temple University Health System, Philadelphia, U.S.	Retrospective cohort study with combined statistical and ML analysis.	n= 67,332,516 (claims data) *sample population not specified.	SEER-Medicare claims database (1992-2014)	Machine Learning- Pointwise Mutual Information (PMI)
Gao W. et al 2023[53]	University of Maryland, Baltimore, MD, U.S.	Developmental study with DL system design.	Training data=200,000 testing data n= 222	Camelyon Challenge dataset (2016).	Deep Learning- Convolutional Neural Network (CNN)
Park et al. 2023 [54]	University of Pennsylvania, Philadelphia, U.S.	Retrospective cohort study with ML model development.	n= 322 348	SEER database (2000-2017).	Machine Learning- 1.Gradient Boost Tree (GBT), 2.Cox Proportional hazard (Cox PH), 3.Survival support Vector Machine (SSVM), 4.Survival Tree (ST)
Soltan and Washington 2024 [52]	University of Hawaii, Honolulu, U.S.	Retrospective study with DL classification.	n=10,586	Nightingale Open Science dataset (2014-2020)	Deep Learning- 1.Convolutional Neural Network (CNN)+ Post processing fairness

Stone et al. 2023 [49]	Clemson University, Santa Cruz, U.S.	Observational with comparative analysis study.	n= 200	TGCA repository (unknown time period)	Deep Learning-1.Neural Network (NN)

3.3 Potential Biases

3.3.1 Sampling Bias

Among the studies, only Egleston et al.[51] and Park et al.[54] used randomization and matching, enhancing the internal validity of their studies. Park et al.'s approach maintained original race/ethnicity ratios, which is a strength, whereas Egleston et al. omitted some relevant variables like neighborhood-level income, potentially affecting the outcomes. Gao W. et al [53], Soltan and Washington[52], and Stone et al.[49] did not clearly specify the sampling methods, leaving room for unassessed selection bias.

3.3.2 Omission Bias

Most studies addressed incomplete data to improve model consistency, though this could introduce omission bias and reduce the accuracy of the models if the exclusion is not random[48]. Dell'Aquila et al. excluded patients with missing pathological complete response (pCR) outcomes and non-MRI data, which could skew results if the missing data were not randomly distributed[60]. Dong W. et al. excluded Massachusetts data for 2018, but the handling of missing data was not detailed[50]. Egleston et al. removed managed care contract data[51] while Park et al. excluded records with missing values[54]. Also, Stone et al. cropped images to a fixed size, which may affect outcomes based on retained tissues[49], while Soltan and Washington did not specify their handling of missing data which is a significant oversight[52].

3.3.3 Quality of Data and Historical Bias

The performance of AI models is intrinsically dependent on the quality of the data upon which they are trained[48,63]. All studies rely on pre-recorded databases, which can introduce errors due to inconsistent recording practices, human errors, and variations in data quality which equally threaten internal and external validities of the studies[48]. SEER, used in studies like those by Dong W. et al. [50] and Egleston et al.[50] is noted for its data quality[57], whereas single institution databases used by Dell'Aquila et al.[60] may harbor unaddressed inconsistencies, contingent on the rigor of quality control measures within their medical record systems. Image studies face additional challenges, with Stone et al. using less precise image cropping into (1100 x 2011 pixels)[49] compared to higher-resolution (100,000 x 150,000 pixels) methods in Soltan and Washington's study[52].

3.3.4 Longitudinal Data Fallacy and Temporal Bias.

Dong W. et al. attempted to harmonize data across years assuming no significant changes over time[50]. This may not hold true given the evolution of healthcare practices and technology over the extended period during which the study data was collected, consequently affecting the internal validity of the study.

3.3.5 Exposure and Modelling Bias

Biases in model development can equally threaten a study's internal and external validity. Dell'Aquila et al.[60], Park et al.[54], and Soltan and Washington [52] followed standard partitioning ratios (80:20, 70:30, 80:20)[48] enhancing model validation. The lack of detailed partitioning information in other studies limits the assessment of their model's robustness.

3.3.6 Confounding and Predictor Variables

In machine learning models, various feature engineering techniques are used to control for

confounders. It is essential to select an appropriate number of features relative to the sample size to ensure model accuracy and minimize confounding[48]. All studies adhered to a suitable feature-to-sample size ratio (more than 1:10 ratio[64]) but some missed key predictors, potentially introducing confounding. For example, Dell'Aquila et al. excluded comorbidities and lifestyle factors[60], and Egleston et al. did not account for income levels[51], which are significant for the outcomes measured. Park et al. included a comprehensive list of predictors but lacked critical socioeconomic factors[54], while Dong W. et al. included the most predictors, enhancing model accuracy[50].

3.3.7 Outcome Measures

Four studies[50,51,54,60] employed variations of ML algorithms to develop intervention models. Additionally, three studies[49,52,53] utilized DL techniques to process image data. This methodological diversity has resulted in significant heterogeneity in evaluation metrics (accuracy, F1-score, ROC AUC), complicating the direct comparison of outcomes due to differing performance emphases across metrics.

Study findings, strengths, limitations, and bias of each study are summarized in Multimedia Appendix named "Summary of study findings, strengths, limitations, and bias".

3.4 Equity-Related Findings

Due to the heterogeneous nature of the equity rationale and study types, equity-related findings were analyzed thematically. This involved extracting findings from each study and grouping them by common themes. The review identified three main themes regarding AI's impact on equity in BC: forecasting biological and socio-economic factors affecting outcomes, enhancing decision-making fairness, and use of AI to reduce healthcare inequalities in a global context. A summary of the equity-related rationale, objectives, and outcomes is presented in Table 5, with a detailed discussion in Section 4.

Table 5: Summary of the equity-related rationale, objectives and outcomes.

Study	Study Rationale in Relation to Health Equity	Study Objective	Health Equity related findings.
Studies focusing on forecasting biological and socio-economic factors influencing cancer outcomes.			
Dell'Aquila et al.[60]	Predicting treatment and survival outcomes in diverse populations is challenging due to complex interactions between biological and socioeconomic factors.	To employ four ML models to identify predictors of pCR and OS in breast cancer patients and integrate diverse factors for improved predictions.	1. pCR is associated with tumor stage, and tumor size and BPE, but not race, ethnicity, income quintile, and insurance status. 2. OS is associated with pCR status, tumor subtype, tumor stage, some MRI data, and insurance status. 3. There was no association between race/ethnicity and pCR or OS once adjusted for tumor subtype.
Dong W. et al.[50]	Identifying complex combinations of factors contributing to geographical disparities in LSBC diagnosis is challenging. Traditional subjective stratification methods often overlook nuanced interactions between these factors.	To apply ML techniques for analyzing complex interactions among broader determinants, with the aim of identifying county areas ("phenotypes") to predict the risk levels of LSBC diagnosis and map their geographic distribution.	1. pCR is associated with tumor stage, and tumor size and BPE, but not race, ethnicity, income quintile, and insurance status. 2. OS is associated with pCR status, tumor subtype, tumor stage, some MRI data, and insurance status. 3. There was no association between race/ethnicity and pCR or OS once

			adjusted for tumor subtype.
Egleston et al.[51]	Accurate identification of medical procedures from Medicare claims data is challenging due to complex coding systems. Traditional methods may miss treatments and affect research outcomes which could lead to incorrect visualization of racial disparities.	To develop a ML model to accurately identify diagnosis and procedure codes from Medicare claims data which could have been missed earlier, improving understanding of racial disparities in breast cancer survival outcomes.	1. The racial disparities in breast cancer specific survival and overall survival diminished when adjusted for augmented treatment definitions.
Stone et al.[49]	Variations in tumor biology; ECM, which human eye is unable to distinguish, may contribute to disparities in cancer incidence and survival across racial groups.	To investigate racial differences in breast tumor morphologies, ECM regions and tumor regions using DL methods.	1. Significant differences in ECM regions between African American and Caucasian groups were identified. 2. Mixed results on differences of tumor regions between the two groups.
Studies focusing on influencing the fairness of decision-making.			
Park et al.[54]	Traditional ML models often fail to predict outcomes accurately for underrepresented groups due to representation biases in training data.	To develop race/ ethnicity-specific survival ML models for Hispanic and Black women with breast cancer and compare them to general models to determine improved prediction accuracy for	1. Race/ethnicity-specific ML models had improved prediction accuracy compared to general models.

		these populations.	
Soltan and Washington.[52]	AI for cancer diagnostics often faces challenges in achieving fair outcomes across diverse demographic groups due to uneven dataset representation.	To evaluate bias in DL models for breast cancer stage prediction and assess the effectiveness of post-processing methods in mitigating this bias for white and non-white population samples.	1. Without post-processing methods, all models demonstrated consistently higher binary accuracy, precision, and recall for the White group compared to the non-White group. 2. Fairness adjustments did not lead to consistent improvements of non-biased predictions across the two groups.
Studies focusing on use of AI to reduce healthcare inequalities in a global context.			
Gao W. et[53]	High breast cancer mortality rates in developing countries are due to delays in diagnosis caused by a shortage of pathologists and inadequate healthcare infrastructure.	To develop a DL based diagnostic system for metastatic breast cancer that is accurate and efficient, comparing CNN architectures to find the best balance for under-resourced environments.	1. MobileNetV2 was a suitable diagnostic model for low-resource settings, with improved diagnostic accuracy and mobile efficiency.
Abbreviations in the table: pCR= pathological complete response, OS= Overall survival, LSBC=Late-Stage Breast Cancer			

4.Discussion

4.1 Synthesis of Results

Among the identified three principal themes, the predominant focus of the reviewed articles was on predicting the biological and socio-economic factors related to BC outcomes.

Dong W. et al.’s findings provide novel and more detailed insights into the distribution of regional disparities in late-stage BC[50] aligning with the previous findings of Mobley et al. who used traditional statistical methods[65]. While these insights can help guide resource allocation, it’s important to note that the geographical areas identified at the county level may still contain significant internal variations; a need for further stratification to address internal disparities more effectively.

However, Dong W. et al.’s findings highlight insurance status as a key predictor variable, which Dell’Aquila et al. study also validated through four different models[50], providing important insight into systemic disparities of breast cancer. Nonetheless, Dong W. et al.’s findings cannot be generalized to individuals, as doing so would lead to an ecological fallacy.

Dell’Aquila et al.’s research incorporated a broader range of biological factors, such as tumor subtype, size, and benign breast tissue enhancement (BPE), which emerged as more significant predictors than socio-economic factors[60]. While studies utilizing traditional statistical methods have explored the link between biological determinants and BC outcomes[66–70], there has been no direct comparison with socio-economic determinants highlighting the advantage of AI models, which can analyze diverse groups of determinants simultaneously.

However, the results should be interpreted cautiously due to the study's limited inclusion of socio-economic predictors.

Stone et al. further emphasized biological influences, identifying significant differences in the extracellular matrix (ECM) between African American and Caucasian BC patients[49], providing novel insights to the previous literature that has attempted to explain racial disparities in BC according to tumor characteristics[71]. While this research strengthens the case for individualized approaches to breast cancer, the effects of these associations need to be carefully evaluated by controlling for various confounders.

Consistent with Dong W. et al., Dell'Aquila et al., Egleston et al. found that race and ethnicity were not as strong predictors as biological and socio-economic factors in BC disparities, arguing that, survival differences between two patient cohorts can be attributed to missed treatment definitions in traditional analyses, thus reinforcing the findings of Silber et al.[72]. However, the lack of statistical significance in these findings suggests the presence of underlying confounding factors.

Collectively, these studies exemplify AI's capability to analyze large and complex datasets, considering the interplay between numerous predictor variables. Among the studies reviewed, Dong W. et al. offered the most comprehensive analysis and methodological rigor, thereby inspiring greater confidence in their conclusions. However, non-representative databases of other studies[49,51,60] raise concerns about the fairness of the findings.

Two studies examined AI's role in enhancing decision-making fairness, particularly focusing on racial bias introduced by non-representative datasets. These attempts are novel approaches in accompanying wider literature. Park et al.'s model demonstrated superior predictive accuracy, owing to a larger sample size, a more comprehensive dataset, and greater methodological robustness, with a substantial number of controlled predictor variables[54]. Conversely, Soltan and Washington's approach yielded mixed results, partly due to the aggregation of diverse non-White populations into a single category[52]. It is noteworthy that the two studies also analyzed different types of data, further distinguishing their findings.

Nevertheless, both studies confirmed the biased predictive power of general models. Consequently, while AI models trained on biased datasets may exacerbate disparities, race/ethnicity-specific models appear promising in mitigating such biases. However, broader generalization requires models that are trained and tested on diverse populations.

The least explored theme was the use of AI to reduce healthcare inequalities in a global context, with only Gao W. et al addressing this area. Their study highlighted the suitability of AI diagnostic

models for resource-constrained environments in developing countries. This approach addresses previous concerns about implementing AI solutions in regions with limited digital infrastructure[30,36]. However, the study did not address other critical implementation challenges, such as cost-effectiveness or the need for trained personnel, which have been emphasized in earlier literature. Additionally, the development of the model based on datasets from HICs for use in LMICs raises concerns about dataset shifts, potentially exacerbating disparities—a significant ethical issue voiced by Weissglass[73] and Wahl et al[74].

This CLR effectively achieved its goals by thoroughly evaluating the literature and thematically organizing the different ways AI applications can impact disparities in breast cancer. Nonetheless, the review was unable to provide a definitive answer to the research question because of the high variability among studies and the numerous biases present in the included research.

4.2 Strengths and Limitations

This study followed a systematic approach throughout, adhering to the PICOS framework and PRISMA guidelines for the comprehensive search and retrieval of relevant articles. A key strength of this review was the integration of critical appraisal tools from both public health and health informatics disciplines, which provided a robust appraisal of the literature. Additionally, the use of both descriptive and thematic analysis for data extraction enriched the synthesis of the results. These enhanced the methodological rigor and internal validity of the review.

However, the study's scope was limited to peer-reviewed literature, excluding grey literature. This may have introduced a Type I bias by potentially overlooking relevant findings from unpublished or non-peer-reviewed sources. Additionally, the geographical focus on studies from the United States limits the generalizability of the findings, hindering the ability to assess the global impact of AI on health equity. Consistent with earlier literature discussed in the rationale, the studies included were predominantly retrospective, with an absence of superior designs such as prospective or randomized

controlled trials. Furthermore, any methodological biases present in the reviewed studies also represent a limitation of this review.

4.3 Future Policy Implications

A recurring theme throughout this review is the disparities introduced by the inherent bias of non-representative, incomplete, and poor-quality datasets used to train AI models. This issue has been extensively documented in the literature[34,75,76].

National and international cancer databases exhibit distinct challenges concerning representativeness, completeness, and accuracy. The SEER database, the most extensive in the United States, covers only approximately half of the U.S. population, with variable representation across racial and ethnic groups[77]. Moreover, issues of incompleteness and inaccuracies due to unrecorded variables and inconsistencies in coding have been reported[78,79]. Conversely, the UK's National Cancer Registration and Analysis Service (NCRAS) is acclaimed for its thoroughness, precision, and comprehensive recording of cases within the National Health Service (NHS), though some variations and coding issues have been reported[80–82]. The UK's smaller population and the centralized structure of its healthcare system[83], in difference to the U.S.'s more fragmented system[84] may contribute to these differences in database quality.

Current global and national guidelines on data and datasets for AI research and development are still in their infancy, with existing frameworks generally covering AI applications across various sectors collectively rather than focusing specifically on healthcare applications. Recent publications, including the WHO's "Regulatory Considerations on Artificial Intelligence for Health[85]", The European Union's “AI Act”[86], the UK Department for Science and Innovation's whitepaper on AI regulation[87], the FDA's whitepaper on Artificial Intelligence & Medical Products[88], all released between 2023 and 2024, are evident of this nascent stage. These guidelines largely emphasize ethical data governance, focusing on privacy and ethical issues

of handling data, but poorly address the dataset inclusivity and representativeness.

The WHO and EU recommendations primarily focus on dataset management during AI model development rather than improving dataset quality[85,89,90]. This highlights the need for policies that enforce transparency regarding dataset representativeness and quality. Public health policy in AI should investigate methods to enhance dataset inclusivity and establish standards for benchmark datasets used in AI research and development, alongside developing protocols for standardized data collection.

The publication[91] by the UK NHS AI Lab's ethics initiative[92] provides a valuable foundation for creating such frameworks. Developing or adapting similar frameworks should be considered by other countries while the feasibility of integrating this initiative into policy should be explored within the UK context.

Furthermore, in the global context, the struggle of LMICs to establish reliable, complete, and representative health databases is evident from the limitations of quality and coverage of cancer data from some low- and middle-income countries on the Global Cancer Observatory (GCO)[93]. Therefore, it is crucial to address the data poverty in LMICs. Exploring methods to establish databases and surveillance systems that can support unbiased and generalizable AI research in these settings is imperative.

4.4 Recommendations for Future Research

Randomized controlled trials (RCTs) and prospective cohort studies can reveal the true impact of AI on health equity in BC. RCTs are better suited for evaluating AI applications already integrated into BC care while prospective cohort studies are better suited for AI in predictive or classification analysis. This allows the collection of high-quality data, reducing the influence of biased datasets and improving clinical relevance[94]. A conceptual framework for conducting RCTs to evaluate

health equity outcomes has been described by Jull et al.[95]. Significant attention must be paid to the selection and control of predictor variables to minimize confounding biases, which can be achieved through multidisciplinary collaboration[48,96].

Additionally, adherence to standard reporting guidelines is crucial to improve the reproducibility of studies[48,97,98], an aspect that is currently lacking in much of the existing research.

4.5 Conclusion

In conclusion, AI applications in breast cancer diagnosis, treatment, and outcomes have the potential to either mitigate or worsen existing inequities. To minimize this risk, careful attention must be given to the data used and biases in model development. Further evidence is needed to fully understand their true impact.

5. Acknowledgments

We gratefully acknowledge Yiqi Liu, a PhD student of Computer Science from the Department of Computer Science at the University of Manchester [ORCID: 0000-0002-8070-5056], for their generous assistance with proofreading and reviewing the critical appraisal of the studies using the ROBUST-ML checklist during the review process. No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

6. Conflicts of Interest

None declared

7. Data Availability

Data sharing is not applicable to this article as no data sets were generated or analyzed during this study.

8. Abbreviations

1. Breast Cancer (BC)
2. High Income Countries (HICs)
3. Low-Middle-Income Countries (LMICs)
4. Artificial Intelligence (AI)
5. Machine Learning (ML)
6. Deep Learning (DL)
7. United Kingdom (UK)
8. Critical Literature Review (CLR)
9. United States (U.S.)
10. Surveillance, Epidemiology and End Results registries -(SEER)
11. Randomized controlled trials (RCTs)

9. Multimedia Appendix

1. Full Search Strategy

2.Summary of study findings, strengths, limitations, and bias

3.Public Health Critical Appraisal Checklist

4. The Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML) checklist

10. References

1. American Cancer Society. What Is Breast Cancer? Available from: https://www.cancer.org/cancer/types/breast-cancer/about/what-is-breast-cancer.html [accessed Sept 11, 2024]

2. World Health Organization. Breast cancer. Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer [accessed Sept 3, 2024]

3. Barrios CH. Global challenges in breast cancer detection and treatment. Breast Off J Eur Soc Mastology 2022 Feb 8;62(Suppl 1):S3–S6. PMID:35219542

4. Yedjou CG, Sims JN, Miele L, Noubissi F, Lowe L, Fonseca DD, Alo RA, Payton M, Tchounwou PB. Health and Racial Disparity in Breast Cancer. Adv Exp Med Biol 2019;1152:31–49. PMID:31456178

5. Wilson J, Sule AA. Disparity in Early Detection of Breast Cancer. StatPearls Treasure Island (FL): StatPearls Publishing; 2024. PMID:33231981

6. Gehlert S, Hudson D, Sacks T. A Critical Theoretical Approach to Cancer Disparities: Breast Cancer and the Social Determinants of Health. Front Public Health Frontiers; 2021 May 21;9. doi: 10.3389/fpubh.2021.674736

7. Mootz A, Arjmandi F, Dogan BE, Evans WP. Health Care Disparities in Breast Cancer: The Economics of Access to Screening, Diagnosis, and Treatment. J Breast Imaging 2020 Nov 21;2(6):524–529. doi: 10.1093/jbi/wbaa093

8. Hill HE, Schiemann WP, Varadan V. Understanding breast cancer disparities—a multi-scale challenge. Ann Transl Med 2020 July;8(14):906. PMID:32793750

9. Martini R, Newman L, Davis M. Breast cancer disparities in outcomes; unmasking biological determinants associated with racial and genetic diversity. Clin Exp Metastasis 2022 Feb 1;39(1):7–14. doi: 10.1007/s10585-021-10087-x

10. Breast Cancer Research Foundation. Understanding Breast Cancer Racial Disparities. Breast Cancer Res Found. 2023. Available from: https://www.bcrf.org/blog/understanding-breast-cancer-racial-disparities/ [accessed Sept 11, 2024]

11. The World Health Organization. Breast cancer inequities. Available from: https://www.who.int/initiatives/global-breast-cancer-initiative/breast-cancer-inequities [accessed Sept 11, 2024]

12. IBM. What Is Artificial Intelligence (AI)? 2024. Available from: https://www.ibm.com/topics/artificial-intelligence [accessed Sept 11, 2024]

13. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Fam Med Prim Care 2019 July;8(7):2328–2331. PMID:31463251

14. Al Kuwaiti A, Nazer K, Al-Reedy A, Al-Shehri S, Al-Muhanna A, Subbarayalu AV, Al Muhanna D, Al-Muhanna FA. A Review of the Role of Artificial Intelligence in Healthcare. J Pers Med Multidisciplinary Digital Publishing Institute; 2023 June;13(6):951. doi: 10.3390/jpm13060951

15. Farina E, Nabhen JJ, Dacoregio MI, Batalini F, Moraes FY. An overview of artificial intelligence in oncology. Future Sci OA 8(4):FSO787. PMID:35369274

16. National Cancer Institute. Artificial Intelligence (AI) and Cancer. 2024. Available from: https://www.cancer.gov/research/infrastructure/artificial-intelligence [accessed Sept 11, 2024]

17. Godfrey I. Harnessing the power of AI to improve outcomes for people with breast cancer. Inst Cancer Res. Available from: https://www.icr.ac.uk/blogs/science-talk/page-details/harnessing-

the-power-of-ai-to-improve-outcomes-for-people-with-breast-cancer [accessed Sept 11, 2024]

18. Ahn JS, Shin S, Yang S-A, Park EK, Kim KH, Cho SI, Ock C-Y, Kim S. Artificial Intelligence in Breast Cancer Diagnosis and Personalized Medicine. J Breast Cancer 2023 Oct 17;26(5):405–435. PMID:37926067

19. Zheng D, He X, Jing J. Overview of Artificial Intelligence in Breast Cancer Medical Imaging. J Clin Med 2023 Jan 4;12(2):419. PMID:36675348

20. Lauritzen AD, Lillholm M, Lynge E, Nielsen M, Karssemeijer N, Vejborg I, Moy L. Early Indicators of the Impact of Using AI in Mammography Screening for Breast Cancer. Radiology Radiological Society of North America; 2024 June;311(3):e232479. doi: 10.1148/radiol.232479

21. Arasu VA, Habel LA, Achacoso NS, Buist DSM, Cord JB, Esserman LJ, Hylton NM, Glymour MM, Kornak J, Kushi LH, Lewis DA, Liu VX, Lydon CM, Miglioretti DL, Navarro DA, Pu A, Shen L, Sieh W, Yoon H-C, Lee C. Comparison of Mammography AI Algorithms with a Clinical Risk Model for 5-year Breast Cancer Risk Prediction: An Observational Study. Radiology Radiological Society of North America; 2023 June;307(5):e222733. doi: 10.1148/radiol.222733

22. World Health Organization. Health equity. Available from: https://www.who.int/health-topics/health-equity [accessed Sept 11, 2024]

23. Centers for Disease Control and Prevention. What is Health Equity? Health Equity. 2024. Available from: https://www.cdc.gov/health-equity/what-is/index.html [accessed Sept 11, 2024]

24. pwc UK. No longer science fiction, AI and robotics are transforming healthcare. PwC. Available from: https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/transforming-healthcare.html [accessed Sept 13, 2024]

25. NHS England. Artificial Intelligence. NHS Transform Dir. 2023. Available from: https://transform.england.nhs.uk/information-governance/guidance/artificial-intelligence/ [accessed Sept 13, 2024]

26. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 June;6(2):94–98. PMID:31363513

27. Hansen D, Coiera E, Verspoor K. AI is already being used in healthcare. But not all of it is ‘medical grade.’ The Conversation. 2023. Available from: http://theconversation.com/ai-is-already-being-used-in-healthcare-but-not-all-of-it-is-medical-grade-207912 [accessed Sept 13, 2024]

28. Goldenberg K. The Quiet Revolution: How AI is Changing Cancer Medicine. Available from: https://winshipcancer.emory.edu/magazine/issues/2024/spring/features/how-ai-is-changing-cancer-medicine [accessed Sept 13, 2024]

29. Green BL, Murphy A, Robinson E. Accelerating health disparities research with artificial intelligence. Front Digit Health Frontiers; 2024 Jan 23;6. doi: 10.3389/fdgth.2024.1330160

30. Istasy P, Lee WS, Iansavichene A, Upshur R, Gyawali B, Burkell J, Sadikovic B, Lazo-Langner A, Chin-Yee B. The Impact of Artificial Intelligence on Health Equity in Oncology: Scoping

Review. J Med Internet Res 2022 Nov 1;24(11):e39748. PMID:36005841

31. Mema E., McGinty G. The Role of Artificial Intelligence in Understanding and Addressing Disparities in Breast Cancer Outcomes. Curr Breast Cancer Rep United States: Springer; 2020;12(3):168–174. doi: 10.1007/s12609-020-00368-x

32. Viswanathan VS, Parmar V, Madabhushi A. Towards equitable AI in oncology. Nat Rev Clin Oncol Nature Publishing Group; 2024 Aug;21(8):628–637. doi: 10.1038/s41571-024-00909-8

33. Paulus J, Kent D. Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. NPJ Digit Med 2020 July 30;3(1). doi: 10.1038/s41746-020-0304-9

34. Dankwa-Mullan I, Weeraratne D. Artificial Intelligence and Machine Learning Technologies in Cancer Care: Addressing Disparities, Bias, and Data Diversity. Cancer Discov United States: American Association for Cancer Research Inc.; 2022;12(6):1423–1427. doi: 10.1158/2159-8290.CD-22-0373

35. O'Brien N, Van Dael J, Clarke J, Gardner C, O'Shaughnessy J, Darzi A, Ghafur S. Addressing Racial and Ethnic Inequity in Data-driven Health Technologies. Available from: https://ora.ox.ac.uk/objects/uuid:76cf09cf-ddf7-43ae-a724-9b447136356a/files/ssn009z94p [accessed Sept 13, 2024]

36. Yu L, Zhai X. Use of artificial intelligence to address health disparities in low- and middle-income countries: a thematic analysis of ethical issues. Public Health 2024 Sept 1;234:77–83. doi: 10.1016/j.puhe.2024.05.029

37. Elsevier Author Services. Levels of evidence in research. Elsevier Author Serv - Artic. 2021. Available from: https://scientific-publishing.webshop.elsevier.com/research-process/levels-of-evidence-in-research/ [accessed Sept 12, 2024]

38. Chua IS, Gaziel-Yablowitz M, Korach ZT, Kehl KL, Levitan NA, Arriaga YE, Jackson GP, Bates DW, Hassett M. Artificial intelligence in oncology: Path to implementation. Cancer Med 2021 May 7;10(12):4138–4149. PMID:33960708

39. Luchini C, Pea A, Scarpa A. Artificial intelligence in oncology: current applications and future perspectives. Br J Cancer Nature Publishing Group; 2022 Jan;126(1):4–9. doi: 10.1038/s41416-021-01633-1

40. PRISMA. NEW Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA). Available from: https://www.prisma-statement.org/ [accessed Sept 3, 2024]

41. CASP. How to use the PICO Framework to Aid Critical Appraisal. CASP - Crit Apprais Ski Programme. Available from: https://casp-uk.net/pico-framework/ [accessed Sept 12, 2024]

42. Cedars Sinai. AI's Ascendance in Medicine: A Timeline. Cedars-Sinai. Available from: https://www.cedars-sinai.org/discoveries/ai-ascendance-in-medicine.html [accessed Sept 13, 2024]

43. Waikar S. How Has COVID Affected the AI Economy? 2021. Available from: https://hai.stanford.edu/news/how-has-covid-affected-ai-economy [accessed Sept 3, 2024]

44. Kumar A, Gadag S, Nayak UY. The Beginning of a New Era: Artificial Intelligence in Healthcare. Adv Pharm Bull 2021 June;11(3):414–425. PMID:34513616

45. Suarez A, Bisschops J. Breast Cancer Screening Recommendations for Transgender Patients: A Review - Mayo Clinic Proceedings. Available from: https://www.mayoclinicproceedings.org/article/S0025-6196(23)00607-9/fulltext [accessed Sept 3, 2024]

46. American Cancer Society. Breast Cancer Statistics | How Common Is Breast Cancer? Available from: https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html [accessed Sept 13, 2024]

47. Heller RF, Verma A, Gemmell I, Harrison R, Hart J, Edwards R. Critical appraisal for public health: A new checklist. Public Health 2008 Jan 1;122(1):92–98. doi: 10.1016/j.puhe.2007.04.012

48. Al-Zaiti SS, Alghwiri AA, Hu X, Clermont G, Peace A, Macfarlane P, Bond R. A clinician's guide to understanding and critically appraising machine learning studies: a checklist for Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML). Eur Heart J Digit Health 2022 Apr 12;3(2):125–140. PMID:36713011

49. Stone A, Kalahiki C, Li L, Hubig N, Iuricich F, Dunn H. Evaluation of breast tumor morphologies from African American and Caucasian patients. Comput Struct Biotechnol J 2023;21:3459–3465. doi: 10.1016/j.csbj.2023.06.019

50. Dong W., Bensken W.P., Kim U., Rose J., Berger N.A., Koroukian S.M. Phenotype Discovery and Geographic Disparities of Late-Stage Breast Cancer Diagnosis across U.S. Counties: A Machine Learning Approach. Cancer Epidemiol Biomarkers Prev United States: American Association for Cancer Research Inc.; 2022;31(1):66–76. doi: 10.1158/1055-9965.EPI-21-0838

51. Egleston B, Chanda A, Bai T, Fang C, Bleicher R, Vucetic S. Using Pointwise Mutual Information for Breast Cancer Health Disparities Research With SEER-Medicare Claims. Methodol-Eur J Res METHODS Behav Soc Sci 2023 Mar;19(1):43–59. doi: 10.5964/meth.8535

52. Soltan A, Washington P. Challenges in Reducing Bias Using Post-Processing Fairness for Breast Cancer Stage Classification with Deep Learning. ALGORITHMS 2024 Apr;17(4). doi: 10.3390/a17040141

53. Gao W., Wang D., Huang Y. Designing a Deep Learning-Driven Resource-Efficient Diagnostic System for Metastatic Breast Cancer: Reducing Long Delays of Clinical Diagnosis and Improving Patient Survival in Developing Countries. Cancer Inform United Kingdom: SAGE Publications Ltd; 2023;22((Gao, Huang) Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD, United States). doi: 10.1177/11769351231214446

54. Park J, Bozkurt S, Park J, Lee S. Evaluation of race/ethnicity-specific survival machine learning models for Hispanic and Black patients with breast cancer. BMJ Health CARE Inform 2023 Jan;30(1). doi: 10.1136/bmjhci-2022-100666

55. National Cancer Institute. SEER-Medicare Linked Data Resource. Available from:

https://healthcaredelivery.cancer.gov/seermedicare/ [accessed Sept 12, 2024]

56. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. SEER. Available from: https://seer.cancer.gov/index.html [accessed Sept 3, 2024]

57. Duggan MA, Anderson WF, Altekruse S, Penberthy L, Sherman ME. The Surveillance, Epidemiology and End Results (SEER) Program and Pathology: Towards Strengthening the Critical Relationship. Am J Surg Pathol 2016 Dec;40(12):e94–e102. PMID:27740970

58. National Cancer Institute. The Cancer Genome Atlas Program (TCGA). 2022. Available from: https://www.cancer.gov/ccg/research/genome-sequencing/tcga [accessed Sept 12, 2024]

59. National Cancer Institute. Outcomes & Impact of The Cancer Genome Atlas. 2019. Available from: https://www.cancer.gov/ccg/research/genome-sequencing/tcga/history [accessed Sept 12, 2024]

60. Dell'Aquila K., Vadlamani A., Maldjian T., Fineberg S., Eligulashvili A., Chung J., Adam R., Hodges L., Hou W., Makower D., Duong T.Q. Machine learning prediction of pathological complete response and overall survival of breast cancer patients in an underserved inner-city population. Breast Cancer Res United Kingdom: BioMed Central Ltd; 2024;26(1):7. doi: 10.1186/s13058-023-01762-w

61. Grand Challenge. CAMELYON16-Data. Gd-Challengeorg. Available from: https://camelyon16.grand-challenge.org/Data/ [accessed Sept 12, 2024]

62. Nightingale Open Science. Nightingale Open Science - Open data for healthcare AI research. Available from: https://www.ngsci.org/ [accessed Sept 12, 2024]

63. Miller K. Data-Centric AI: AI Models Are Only as Good as Their Data Pipeline. 2022. Available from: https://hai.stanford.edu/news/data-centric-ai-ai-models-are-only-good-their-data-pipeline [accessed Sept 12, 2024]

64. Stupak T. How Much Data Is Required To Train ML Models in 2024? Akkio. Available from: https://www.akkio.com/post/how-much-data-is-required-to-train-ml [accessed Sept 12, 2024]

65. Mobley LR, Tangka FKL, Berkowitz Z, Miller J, Hall IJ, Wu M, Sabatino SA. Geographic Disparities in Late-Stage Breast Cancer Diagnosis Rates and Their Persistence Over Time. J Womens Health 2021 June 1;30(6):807–815. PMID:33926216

66. Liu Y, He M, Zuo W-J, Hao S, Wang Z-H, Shao Z-M. Tumor Size Still Impacts Prognosis in Breast Cancer With Extensive Nodal Involvement. Front Oncol 2021 Apr 9;11:585613. PMID:33898305

67. Rella R, Contegiacomo A, Bufi E, Mercogliano S, Belli P, Manfredi R. Background parenchymal enhancement and breast cancer: a review of the emerging evidences about its potential use as imaging biomarker. Br J Radiol 2021 Mar;94(1119):20200630. PMID:33035073

68. Orrantia-Borunda E, Anchondo-Nuñez P, Acuña-Aguilar LE, Gómez-Valles FO, Ramírez-Valdespino CA. Subtypes of Breast Cancer. In: Mayrovitz HN, editor. Breast Cancer Brisbane (AU): Exon Publications; 2022. PMID:36122153ISBN:978-0-6453320-3-2

69. Haque R, Ahmed SA, Inzhakova G, Shi J, Avila C, Polikoff J, Bernstein L, Enger SM, Press MF. Impact of Breast Cancer Subtypes and Treatment on Survival: An Analysis Spanning Two Decades. Cancer Epidemiol Biomarkers Prev 2012 Oct 7;21(10):1848–1855. doi: 10.1158/1055-9965.EPI-12-0474

70. Narod SA. Tumour size predicts long-term survival among women with lymph node-positive breast cancer. Curr Oncol 2012 Oct;19(5):249–253. PMID:23144572

71. Prakash O, Hossain F, Danos D, Lassak A, Scribner R, Miele L. Racial Disparities in Triple Negative Breast Cancer: A Review of the Role of Biologic and Non-biologic Factors. Front Public Health Frontiers; 2020 Dec 22;8. doi: 10.3389/fpubh.2020.576964

72. Silber JH, Rosenbaum PR, Clark AS, Giantonio BJ, Ross RN, Teng Y, Wang M, Niknam BA, Ludwig JM, Wang W, Even-Shoshan O, Fox KR. Characteristics Associated With Differences in Survival Among Black and White Women With Breast Cancer. JAMA 2013 July 24;310(4):389. doi: 10.1001/jama.2013.8272

73. Weissglass DE. Contextual bias, the democratization of healthcare, and medical artificial intelligence in low- and middle-income countries. Bioethics 2022;36(2):201–209. doi: 10.1111/bioe.12927

74. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health BMJ Specialist Journals; 2018 Aug 1;3(4):e000798. PMID:30233828

75. Arora A, Alderman JE, Palmer J, Ganapathi S, Laws E, McCradden MD, Oakden-Rayner L, Pfohl SR, Ghassemi M, McKay F, Treanor D, Rostamzadeh N, Mateen B, Gath J, Adebajo AO, Kuku S, Matin R, Heller K, Sapey E, Sebire NJ, Cole-Lewis H, Calvert M, Denniston A, Liu X. The value of standards for health datasets in artificial intelligence-based applications. Nat Med 2023;29(11):2929–2938. PMID:37884627

76. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. Npj Digit Med Nature Publishing Group; 2023 June 14;6(1):1–3. doi: 10.1038/s41746-023-00858-z

77. National Cancer Institute. Number of Persons by Race and Hispanic Ethnicity for SEER Participants - SEER Registries. Available from: https://seer.cancer.gov/registries/data.html [accessed Sept 8, 2024]

78. National Cancer Institute. SEER Treatment Data Limitations (November 2022 Submission) - SEER Data & Software. SEER. Available from: https://seer.cancer.gov/data-software/documentation/seerstat/nov2022/treatment-limitations-nov2022.html [accessed Sept 13, 2024]

79. Jairam V, Park HS. Strengths and limitations of large databases in lung cancer radiation oncology research. Transl Lung Cancer Res AME Publishing Company; 2019 Sept;8(Suppl 2). doi: 10.21037/tlcr.2019.05.06

80. National Disease Registration Service. Cancer. NDRS. Available from: https://digital.nhs.uk/ndrs/about/ncras [accessed Sept 13, 2024]

81. Henson KE, Elliss-Brookes L, Coupland VH, Payne E, Vernon S, Rous B, Rashbass J. Data Resource Profile: National Cancer Registration Dataset in England. Int J Epidemiol 2020 Feb;49(1):16–16h. PMID:31120104

82. NHS England. Cancer registration data, quality and methodology. NHS Engl Digit. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/cancer-registration-statistics/england-2019/cancer-registration-data-quality-and-methodology [accessed Sept 13, 2024]

83. Anderson M, Pitchforth E, Edwards N, Alderwick H, McGuire A, Mossialos E, Hernández-Quevedo C. United Kingdom: health system summary. Available from: https://eurohealthobservatory.who.int/publications/i/united-kingdom-health-system-summary [accessed Sept 13, 2024]

84. ISPOR. US Healthcare System Overview-Background. Ispor.org. Available from: https://www.ispor.org/heor-resources/more-heor-resources/us-healthcare-system-overview/us-healthcare-system-overview-background-page-1 [accessed Sept 13, 2024]

85. World Health Organization. key regulatory considerations on artificial intelligence (AI) for health. Available from: https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf [accessed Sept 13, 2024]

86. EU Artificial Intelligence Act. The AI Act Explorer. Available from: https://artificialintelligenceact.eu/ai-act-explorer/ [accessed Sept 13, 2024]

87. Department for Science innovation & Technology. A pro-innovation approach to AI regulation: government response. GOVUK. Available from: https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response [accessed Sept 13, 2024]

88. King & Spalding. FDA Publishes White Paper on Artificial Intelligence & Medical Products.

89. Flowers T, Wu P, Lipstaite S. The EU AI Act: Implications for the health sector. Access Partnersh. 2024. Available from: https://accesspartnership.com/the-eu-ai-act-implications-for-the-health-sector/ [accessed Sept 13, 2024]

90. Vollers N, Dennis A. AI regulation in healthcare: UK and EU approaches. 2023. Available from: https://www.taylorwessing.com/en/insights-and-events/insights/2023/03/ai-regulation-in-healthcare-uk-and-eu-approaches [accessed Sept 13, 2024]

91. The STANDING Together collaboration. Recommendations for Diversity, Inclusivity, and Generalisability in Artificial Intelligence Health technologies and Health Datasets. Zenodo; 2023 Oct 30; doi: 10.5281/zenodo.10048356

92. NHS England. The AI Ethics Initiative. NHS Transform Dir. Available from: https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/ethics/ [accessed Sept 13, 2024]

93. The International Agency for Research on Cancer (IARC). Global Cancer Observatory. Available from: https://gco.iarc.fr/ [accessed Sept 13, 2024]

94. Setia MS. Methodology Series Module 1: Cohort Studies. Indian J Dermatol 2016;61(1):21–25.

PMID:26955090

95. Jull J, Whitehead M, Petticrew M, Kristjansson E, Gough D, Petkovic J, Volmink J, Weijer C, Taljaard M, Edwards S, Mbuagbaw L, Cookson R, McGowan J, Lyddiatt A, Boyer Y, Cuervo LG, Armstrong R, White H, Yoganathan M, Pantoja T, Shea B, Pottie K, Norheim O, Baird S, Robberstad B, Sommerfelt H, Asada Y, Wells G, Tugwell P, Welch V. When is a randomised controlled trial health equity relevant? Development and validation of a conceptual framework. BMJ Open 2017 Sept 25;7(9):e015815. PMID:28951402

96. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J 2021 July;8(2):e188–e194. PMID:34286183

97. Campbell JP, Lee AY, Abràmoff M, Keane PA, Ting DSW, Lum F, Chiang MF. Reporting Guidelines for Artificial Intelligence in Medical Research. Ophthalmology 2020 Dec;127(12):1596–1599. PMID:32920029

98. Shelmerdine SC, Arthurs OJ, Denniston A, Sebire NJ. Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare. BMJ Health Care Inform 2021 Aug 23;28(1):e100385. PMID:34426417

Supplementary Files

Multimedia Appendixes

Full search strategy.
URL: http://asset.jmir.pub/assets/cc15a81a51302bf2c65258d8f2b25f35.doc

Summary of study findings strengths, limitations, and bias.
URL: http://asset.jmir.pub/assets/95b78ee59690179c2cfe5be24a439809.doc

Public health critical appraisal checklist.
URL: http://asset.jmir.pub/assets/4dfa1b491586640230cc7c74edf07e64.doc

The ruling out bias using standard tools in machine learning (ROBUST-ML) checklist.
URL: http://asset.jmir.pub/assets/0e335d7f29631b9c8cf2c86412c61bdc.doc

CONSORT (or other) checklists

PRISMA_2020_checklist.
URL: http://asset.jmir.pub/assets/4d989e57854dc98518f872902d03e953.pdf