Scientific Mapping of Deep Learning Research in Breast Cancer

Jun Liu [1,*] , Shashi Kant Gupta[1]

[1] Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia;

Emails: Jun Liu: pdf.junliu@lincoln.edu.my; Shashi Kant Gupta: shashigupta@lincoln.edu.my

Abstract: Deep learning is reshaping breast cancer research across imaging, pathology, and outcome prediction; however, the structure and recent evolution of this literature have not been comprehensively mapped. This bibliometric review analysed Web of Science Core Collection records published from 2022 to 2026 using CiteSpace 6.4.R1. The keyword network comprised 265 nodes and 1,069 links and resolved into 11 thematic clusters, with mammography occupying a central position. Emerging topics included antibody-drug conjugates, large language models, and prediction of treatment response. The temporal analyses indicate a gradual shift from technical performance evaluation toward multimodal modelling and precision oncology. Future progress will depend on externally validated systems that perform reliably across institutions and patient populations.

Keywords: Breast cancer; artificial intelligence; deep learning; bibliometrics; CiteSpace

1. Introduction

Breast cancer ranks among the most prevalent cancers for women being one of the major causes of deaths associated with cancer [1,2]. It generates many different kinds of data in terms of its clinical management such as mammograms, ultrasounds, MRIs, pathology slides, molecular profiles, and medical records related to treatment. Integrating these heterogeneous data is rather difficult which just emphasizes the role of Artificial Intelligence (AI) especially deep learning technologies.

Deep learning technologies differ from their predecessors in that rather than depend on rules we have painstakingly built, they have the ability to learn only the features relevant to problems from the data. Among the neural networks that we can use are convolutional neural networks (CNN), capable of solving the issues of lesions detection, image classification, regions-of-interest segmentation, breast density estimation, histological grading, and predicting receptor status. Currently, such techniques as transformers, foundation models as well as large language models (LLM) were introduced to literature. Many of the reviews cover the topic of computer-assisted detection systems [3-15], however, they are limited by the focus on a specific modality or task in this area and do not give a full picture of the connections between imaging, pathology, and molecular data.

Another concern is the fact that positive performance reported, including high AUCs and sensitivity, is often derived from retrospective datasets and may not apply to the clinical perspective. Hence, a systematic overview of the field is required to separate topics that seem to be research-worthy from the approaches that are already mature enough for clinical use.

In this way, the readers can comprehend that the purpose of this review is to employ CiteSpace to depict the knowledge domain concerning breast cancer studies regarding AI and deep learning. The objectives include: (1) to understand the conceptual networks of the field; (2) to discover the most important themes and their temporal developments; (3) to find recently used keywords and citation patterns; (4) to relate the bibliometric results to the priorities of research as well as to clinical applications.

2. Related Work

There is already considerable research that demonstrates the use of deep learning in mammogram detection and in risk estimation, in population screening, in ultrasound classifications and lymph nodes evaluations, and in computational pathology with the use of whole slide imaging technology. Radiomics and multi-omics integration have also sparked interest. Some of the most well-known software such as CiteSpace and Kleinberg's burst detection has been around for some time, and researchers can rely on them to find trends in science. However, most of the reviews out there do not give a complete description and mention only one modality. A map that connects all imaging, pathology, biomarkers, and treatment-related research does not exist.

3. Materials and Methods

3.1 Data source and search strategy

We queried the Web of Science Core Collection (WoSCC) using a topic search that combined disease and technology terms:

TS = (("breast cancer" OR "breast neoplasm*" OR "breast carcinoma") AND ("artificial intelligence" OR "deep learning"))

The CiteSpace analyses covered publications from 2022 to 2026, and the search was conducted on 20 July 2026. To ensure reproducibility, the final manuscript should specify the WoSCC editions searched, language restrictions, eligible document types, and the number of records before and after deduplication. This information may be reported as follows: “The search was conducted on [date] across [editions]. Only [document types] published in English between 2022 and 2026 were considered. After removing duplicates, [X] records were retained for analysis.”

The search strategy was intentionally specific because it required an explicit reference to AI or deep learning. Consequently, it may have omitted relevant studies indexed only under terms such as machine learning, neural networks, radiomics, computer-aided diagnosis, transformers, or foundation models. This limitation is considered further below.

3.2 Eligibility and data processing

Full records and cited references were downloaded in plain-text format. Ideally, records should be screened for relevance and duplicates removed before analysis, with a PRISMA flow diagram documenting eligibility criteria, exclusions, and independent duplicate screening. In the present study, we analysed the supplied dataset as provided and focused on interpretation of the resulting maps.

3.3 CiteSpace configuration

All visualizations were generated using CiteSpace 6.4.R1 Advanced. The analysis covered 2022–2026 in one-year slices, used the g-index (k = 25) for node selection, a link-retaining factor of 2.5, a 10-year look-back window, and Pathfinder pruning. The resulting keyword network contained 265 nodes and 1,069 links (density = 0.0306).In the CiteSpace maps, larger nodes generally indicate greater keyword frequency or citation counts. Coloured rings represent different time slices (warm colours denote earlier slices and cool colours later slices), while purple rings indicate high betweenness centrality and may identify nodes bridging distinct parts of the network. Links represent co-occurrence or co-citation rather than causal relationships. Cluster quality was assessed using modularity Q and weighted mean silhouette S.

3.4 Analytical framework

Six outputs were examined: keyword co-occurrence, keyword clustering, cluster timelines, cluster time-zone evolution, keyword citation bursts, and reference citation bursts. These quantitative outputs provided the analytical structure, which was interpreted in light of representative publications within each cluster. As a bibliometric study, this review did not include formal statistical testing or risk-of-bias assessment.

4. Results

4.1 Key Contribution and overall keyword structure

This article provides an overview of 11 thematic communities in breast cancer AI research through analysis done using CiteSpace. It highlights the differentiation of established imaging applications from the emerging precision-oncology practices, the trend of the subject area moving beyond technical accuracy and discussing patient outcome, and uses bibliometric indicators to identify emerging frontiers. It also emphasizes the methodological issues that should be resolved before the wide usage of these models in practice.

As illustrated in Figure 1, the keyword co-occurrence network is highly compact in nature. The bigger nodes are indicative of the major diseases and other key concepts like breast tumors, breast cancer, machine intelligence, deep learning, artificial intelligence, and neural networks. The secondary themes in this regard include mammogram, ultrasound, digital pathology, magnetic resonance imaging, segmentation, and classification. New concepts, such as large language models, conjugated antibodies, and precision medicine, exist as smaller nodes, signifying that these are ideas which are still emerging.

The network is essentially made up of three distinct layers that are closely connected: a tech layer comprised of ML, DL, CNNs, classification, and segmentation; a clinical data layer that is made up of mammography, ultrasound, MRIs, digital pathology, expression information, and biomarkers; and an emerging outcome layer that includes full pathological response, neoadjuvant therapies, chemoresistance, and LLMs. Their overlapping layer indicates a considerable interest in correlating computational methods with clinical endpoints.

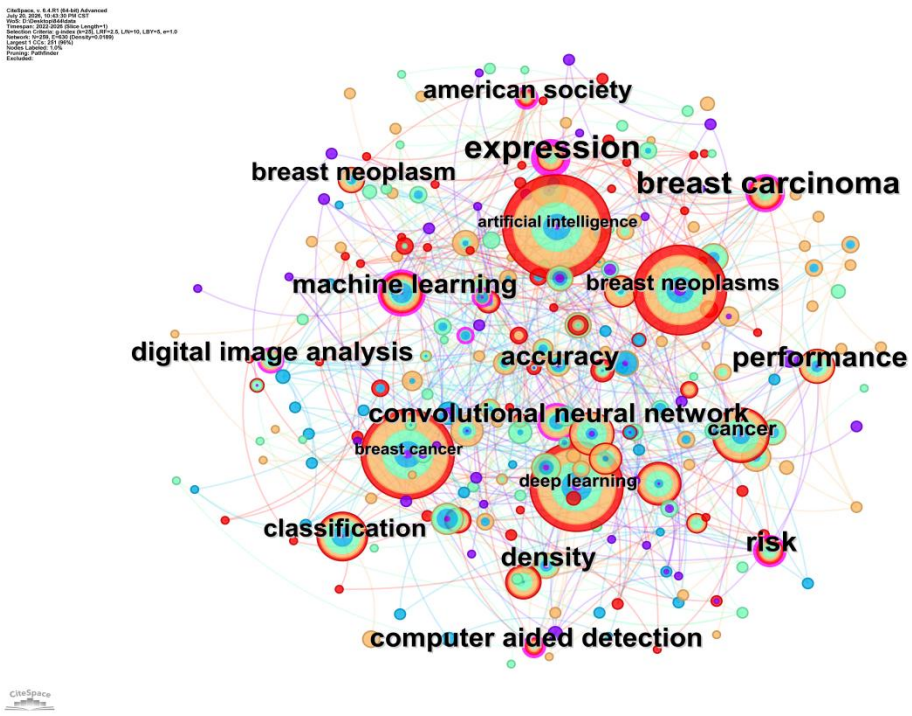


Figure 1. Keyword co-occurrence network of breast cancer research associated with artificial intelligence and deep learning, 2022–2026. The network contains 265 nodes and 1,069 links (density = 0.0306). Larger nodes represent more frequent keywords, and links represent co-occurrence relations. The map highlights breast neoplasms, artificial intelligence, deep learning, machine learning, convolutional neural networks, digital image analysis, computer-aided detection, classification, accuracy, breast density, risk, and performance.

4.2 Thematic clusters

The cluster analysis (Figure 2) showed a modularity Q of 0.6064 and a weighted mean silhouette S of 0.8168 which demonstrates a clear thematic structure and a high degree of internal cluster coherence. A total of eleven clusters were identified: #0 mammography, #1 ultrasonography, #2 breast carcinoma, #3 digital pathology, #4 convolutional neural networks, #5 antibody-drug compounds, #6 carcinoma, #7 breast cancer diagnosis, #8 machine learning, #9 convolutional neural networks, #10 large language models. Clusters #4 and #9 having CNN as part of their names probably indicate semantic differences among their authors discussing related works.

Mammography (#0) constituted the biggest and most central cluster, which corresponds to the scale of screening programmes, presence of extensive image databases, and longevity of computer-aided detection. Ultrasonography (#1) was linked to lesion categorization, elastography, and its complementary role in patients with dense breasts. Digital pathology (#3) was associated with whole-slide imaging, weak learning, and automatic grading.

The antibody-drug conjugate cluster (#5) is remarkable for not being an imaging technique; its existence indicates the emerging connections between computational phenotyping and treatment choices, including HER2 expression, spatial heterogeneity, and prediction of responses. The LLM

cluster (#10) is also quite small and new. Although this information remains uncertain, it indicates the development of new literature synthesis, report generation, and clinical support systems.

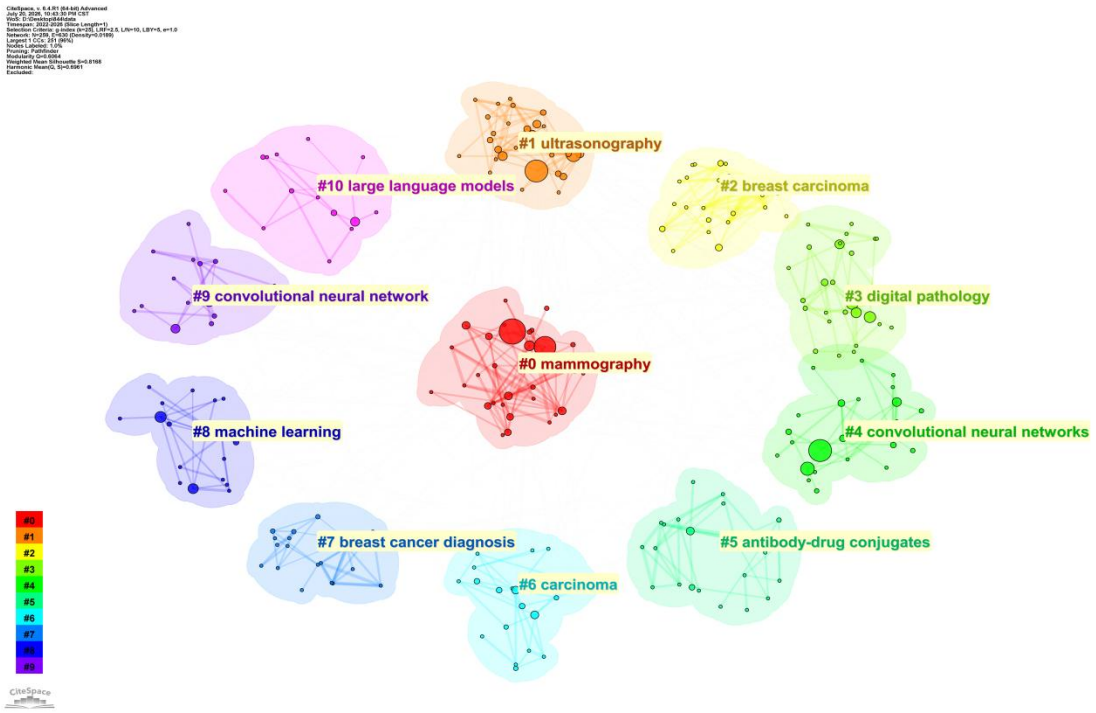


Figure 2. Cluster map of keyword co-occurrence. Eleven clusters were detected: #0 mammography, #1 ultrasonography, #2 breast carcinoma, #3 digital pathology, #4 convolutional neural networks, #5 antibody–drug conjugates, #6 carcinoma, #7 breast cancer diagnosis, #8 machine learning, #9 convolutional neural network, and #10 large language models. Modularity Q was 0.6064 and weighted mean silhouette S was 0.8168.

4.3 Temporal evolution of research themes

The position of various concepts through the years 2022-2026 is given in Figure 3. In mammography, the early studies on effectiveness and breast tumors paved the way for the research of screening mammography, computer aided detection, tomography, detection of tumors, and screening manuals. At the same time ultrasonics have followed the same way as mammography but have maintained a stable focus on diagnosing and machine learning aspects.

The CNN clusters (#4 and #9) included segmentation, classification, lesion features, transfer learning, and algorithm development. The diagnosis and machine-learning clusters extended toward prognosis, prediction, histological grade, distant recurrence, biomarkers, MRI, and neoadjuvant therapy. LLMs and antibody-drug conjugates appeared in the most recent period (2025−2026). Although these remained small nodes, their appearance indicates an expanding research vocabulary.

The time zone graph (Figure 4) offers an insightful observation. Mammography and ultrasound are used the most frequently at the beginning of the study, while breast tumors, digital pathology, antibody-drug conjugates, and LLMs gained prominence later. This trend doesn’t indicate that imaging has lost its relevance; it is indicative of the spread of the field instead.

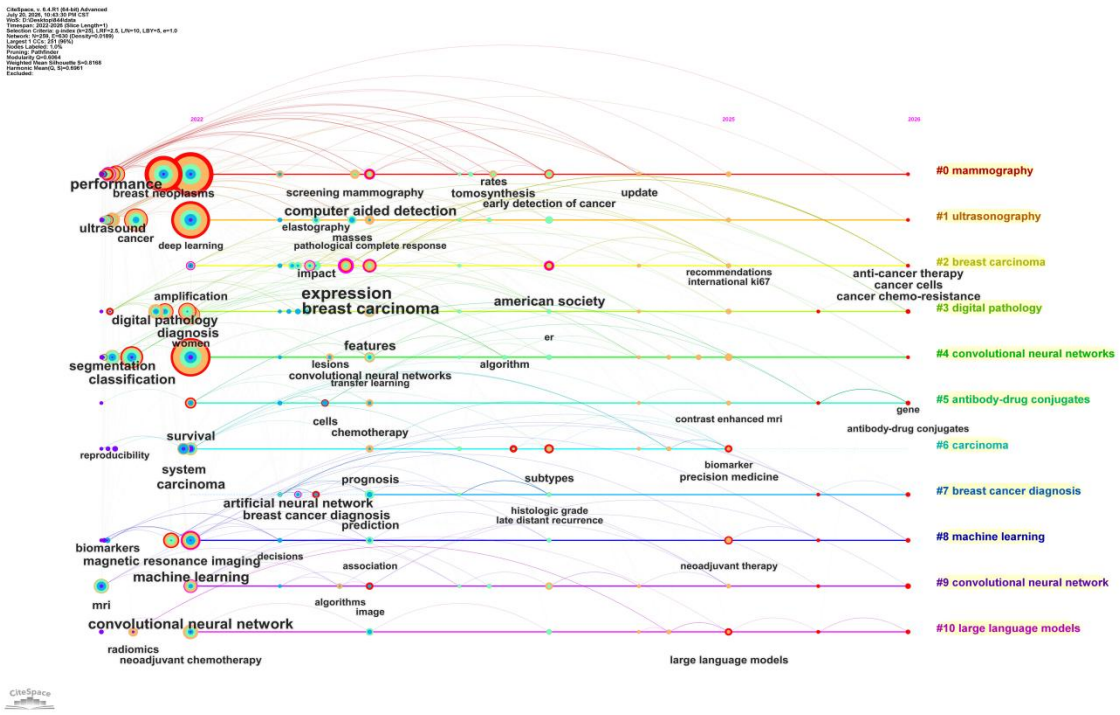


Figure 3. Timeline visualization of keyword clusters, 2022–2026. Nodes are arranged chronologically within thematic clusters. The visualization shows the development from modality- and task-centered research toward biomarkers, prediction, therapy response, precision medicine, and large language models.

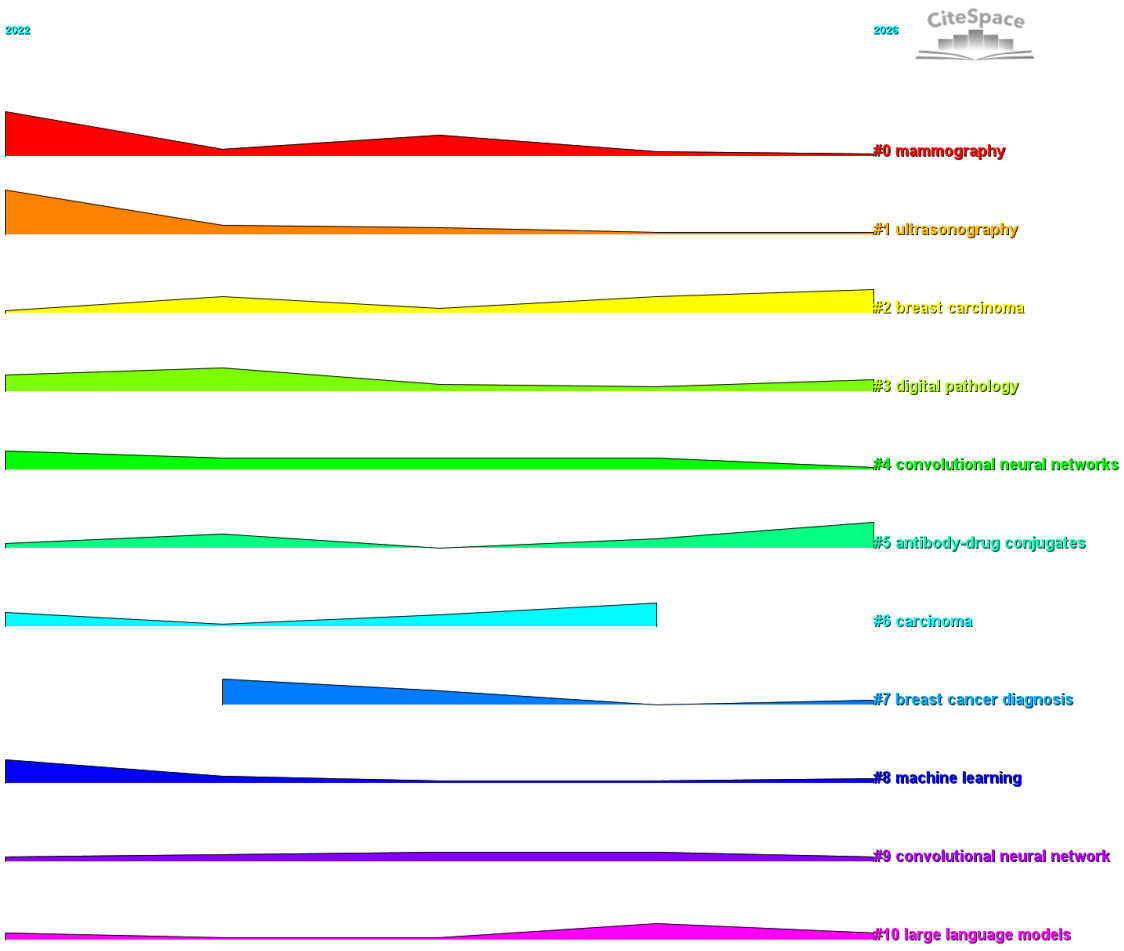


Figure 4. Time-zone visualization of cluster evolution, 2022–2026. The relative prominence of mammography and ultrasonography is greatest early in the period, whereas digital pathology, breast carcinoma, antibody–drug conjugates, and large language models show later or renewed activity. The final year is incomplete and should be interpreted cautiously.

4.4 Keyword citation bursts

The strongest 24 keyword bursts are demonstrated in Figure 5. The strongest burst index for the word ” system” was observed in between 2022−2023 (2.98), followed by ” US” (2.69), ” images” (1.22), ” MRI” (1.19), ” biomarkers” (1.00), and ” breast tomosynthesis” (1.00). Other earlier keyword bursts show continued concern for technical performance and workflows.

The time frame of 2023−2024 demonstrated distinction. “Prediction” had a dominant peak value (1.72) ahead of tumor-infiltrating lymphocytes and pathology (each attaining 1.23), segmentation (1.16), lymph-node metastasis, estrogen, elastography, association, automated image analysis, assay and agreement. In summary, all of these terms imply that the focus has shifted from simply identifying the properties of the models to evaluating their predictability.

Between 2024 to the end of the study, we can find four bursts: “rates” in use, the severed word “outcm” (possibly meant for “outcome” or “outcomes”), “tomosynthesis” (0.75), and

the “algorithm” (0.37). These terms could signify topics of importance that will continue to be of interest. Verification of the truncated term is important before it can be interpreted.

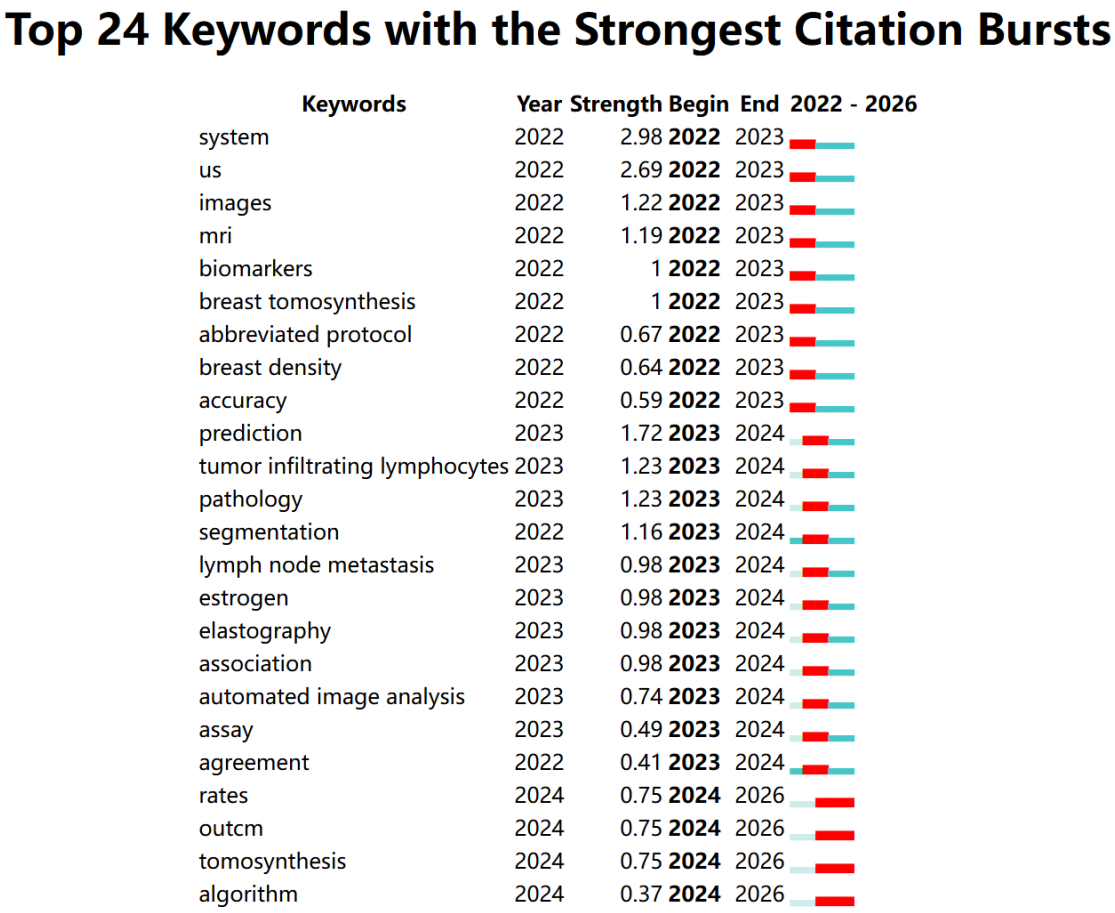


Figure 5. Top 24 keywords with the strongest citation bursts. Red segments show burst intervals. Early bursts include system, US, images, MRI, biomarkers, breast tomosynthesis, abbreviated protocol, breast density, and accuracy. Later bursts include prediction, tumor-infiltrating lymphocytes, pathology, segmentation, lymph-node metastasis, automated image analysis, outcomes-related terminology, tomosynthesis, and algorithm.

4.5 Reference citation bursts and intellectual foundations

Figure 6 shows the 25 strongest reference citation bursts. Couture et al. exhibited the highest burst strength (4.31, 2022–2023); their study linking histological images with grade, receptor status, and molecular features received substantial attention [8]. Other early burst references concerned computational pathology, weakly supervised learning, radiomics, imaging guidelines, and AI reporting checklists [3–5].

The 2023–2024 burst period included studies of mammographic density and risk models [18,19], AI-supported screening [20–22], and consensus reporting recommendations [3]. Bursts beginning in 2024 included deep learning for detection [23], AI in breast imaging [24,25], and foundational natural-language-processing and computer-vision methods supporting recent multimodal

research. Together, these patterns suggest a transition from performance benchmarking toward clinical deployment, standardization, and multimodality.

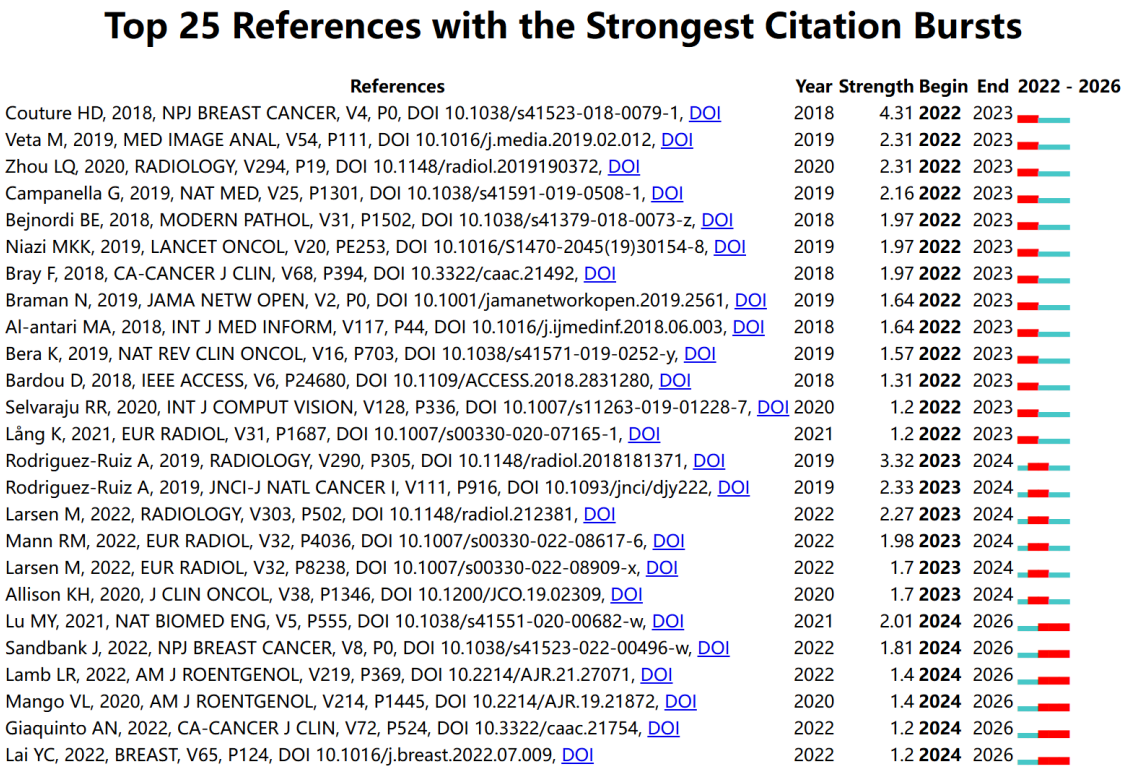


Figure 6. Top 25 references with the strongest citation bursts. Red segments indicate periods of unusually rapid citation growth. The bursts highlight foundational work in computational pathology, weakly supervised learning, radiomics, breast imaging, mammographic risk assessment, reporting guidance, and contemporary breast cancer epidemiology.

5. Discussion

5.1 Current status and challenges of various imaging modalities for breast cancer

In conclusion, the map suggests that breast cancer AI is still mostly reliant on imaging; however, it is gradually becoming multimodal. The foundation of the network is mammography, which links other domains represented by ultrasonography, MRI, digital pathology, and molecular expression. Conventional neural networks still dominate the technical vocabulary, with transformers and large language models being mentioned but not necessarily becoming more prominent than convolutional models in the context of bibliography.
A second approach is the merging of the fields of detection and precision oncology. Words like tumor-infiltrating lymphocytes, lymph-node metastasis, estrogen, Ki-67, biomarkers, pathological complete response, chemoresistance, and antibody-drug conjugates are largely related to biological characterization, response prediction, and therapy guidance. As a result, the research agenda is growing beyond detection into the support of clinical decision making.
One reason that mammography plays a pivotal role in clinical practice is because of the usefulness of the method in providing images to diagnose and screen breast cancer. Several

learning models have been developed to detect lesions, classify cases, estimate density, determine interval-cancer risk, and triage workflow. The introduction of digital breast tomosynthesis for obtaining pseudo-3D mammographic images is expected to change the practice of 2D mammography, but this method also poses challenges of managing data from larger files, reconstruction artifacts, and increased complexity of image annotation. Within a routine context, the results provided by a given model can depend heavily on factors such as the vendor of the equipment, testing techniques used, the age of the patients involved, the extent of breast density, the prevalence of diseases, or screening policies applied. A model trained on rich historical data could provide a satisfactory performance in the retrospectively based testing, but its performance in actual screening can be much worse. More than one reader and one patient with a disease should be involved in studies aimed at validating the performance of the model in different geographical areas.

Ultrasound presents different challenges from mammography. It is operator dependent and images have a naturally high variability, which makes automated standardization attractive but harder to implement. Current work includes mass classification, lesion segmentation, elastography, and prediction of nodal or treatment outcomes. The challenge is that models can have poor transferability between different transducers, scanning protocols and practice environments. Most studies published are single center studies, meaning it is difficult to comment on their generalizability.

Breast MRI holds the potential to provide multiparametric information and high sensitivity, at the expense of its complexity. Applications of AI in this field include lesion detection, segmentation, abbreviated protocols, radiomics, molecular subtyping, treatment response detection, and recurrence risk determination. The models have to deal with enhancement patterns over time, various field strengths, selection of sequences, and the effect of vendors. Single-center studies are the norm here and finding studies that validate the results externally is rare.

Digital pathology has evolved from classification at the patch level to weakly supervised learning focused on full slide images [9,10]. The implementation of multiple-instance learning is beneficial since it allows for the use of whole slide labels during training without the requirement for pixel level labelling. Its applications include the detection of metastasis, grading, determination of receptors and molecular subtypes, and the prediction of treatment outcomes/survival. Still, it remains to be proven through rigorous validation.

There are still significant challenges that we need to overcome. Variations in slide scanners, staining methods, tissue preparation, and types of cases result in a phenomenon called domain shift. Models may utilize artifacts or shortcuts specific to a given site instead of biologically meaningful characteristics. To reliably assess the models one must perform testing across various types of stains, scanners, and institutions, utilize data from individual patients to avoid data leakage between the training and test sets, and validate the models using independent cohorts.

5.2 From prediction to treatment selection

The rise of treatment-oriented terms—including pathological complete response, neoadjuvant therapy, chemoresistance, and antibody-drug conjugates—suggests that AI models are increasingly being developed to support treatment decisions rather than diagnosis alone. Multimodal approaches combining imaging, histology, molecular assays, and clinical variables represent a logical next step for predicting response and recurrence.

But treatment prediction raises the bar. Good discrimination alone is insufficient; calibration also matters because predicted probabilities could influence actual treatment choices. Decision-curve analysis helps here: it asks whether using the model to guide decisions offers net benefit over treating everyone or treating no one, across a range of clinically relevant thresholds. Ideally, models should be benchmarked against current multidisciplinary team decisions or validated clinical scores, not just against other algorithms.

5.3 Large language models and foundation models

Careful evaluation is necessary when working on LLM cluster. The topic is relatively novel and does not have much representation in online media. Possible applications of LLM cluster are information extraction from free-text reports, generation of patient summaries, information synthesis from long-term records, patient matching to clinical studies, etc. Major limitations involve hallucinations, factual inaccuracies, privacy violations and lack of solid evaluation tools in the area.

Though pretraining and the transfer of learning might mitigate the need for annotation, broad pretraining does not guarantee local clinical reliability. Therefore, before deploy their systems in their hospitals, developers need to provide documentation about the training data, overlapping between the training and testing data, adaptation methods they use, and how their models perform in different subgroups.

5.4 Methodological priorities for clinical translation

Our analysis and the wider literature identify five methodological priorities for clinical translation: (1) external validation that reflects variation in institutions, devices, demographics, and disease prevalence; (2) transparent reporting of cohort construction, missing data, reference standards, patient-level splits, and follow-up completeness; (3) prospective evaluation, ideally randomized, for models nearing clinical use; (4) routine fairness analyses across age, ethnicity, socioeconomic status, and breast density; and (5) workflow studies that assess practical outcomes such as reading time, recall rates, biopsy rates, and cancer detection in addition to accuracy.

Reproducibility also deserves explicit attention. Code, or at minimum detailed methods, should be available; models should be versioned; test sets should remain stable; and potential data leakage should be addressed directly. Comparative studies require strong clinical baselines and should report whether gains are statistically as well as clinically meaningful. Federated learning, public benchmarks, and data-sharing consortia could help, although each presents regulatory and logistical challenges.

5.5 Interpretation of bibliometric indicators

Bibliometric indicators require careful interpretation. Big nodes mean frequency or citations, not clinical value. Links are about intellectual association, not causal evidence. Clusters are algorithmic groupings, not natural categories. Bursts reflect attention spikes, not quality. None of these metrics directly measures diagnostic accuracy, bias, evidence quality, or patient benefit. They are useful for characterizing the structure of a field, but they do not establish whether an intervention or model is clinically effective.

The observed Q and S values support the structural validity of the clustering, but cluster labels remain algorithmic summaries. The separation of clusters #4 and #9, both CNN-related, illustrates how lexical variation can divide conceptually similar work. Manual inspection of publications within

each cluster can aid interpretation, although some ambiguity at cluster boundaries is unavoidable in bibliometric analysis.

6. Limitations

There are certain limitations that need to be considered. Firstly, the research is based on WoSCC only, therefore other databases like Scopus, PubMed, IEEE Xplore, arXiv, and conference papers may not be included. Secondly, the requirement of specific terms like AI, deep learning, etc. might have left out some relevant studies that only use terms like machine learning, neural networks, etc. Thirdly, bibliometric tools may not give enough importance to studies that are based in the global south or other non-English languages because these tools are biased towards the studies that are popular, trusted, and published in English. Fourthly, CiteSpace parameters such as g-index, pruning, and time slicing impact the final maps. Finally, the last year must be treated differently. The metrics were computed on July 20, 2026 with 2026 as the last year involved, hence making this time slicing incomplete. If there were statistics computed prior to this date or if there were any future data, the same process for development of data and metrics must be repeated one more time. The date of research must be stated clearly in the paper.

7. Future Directions

In the future, bibliometric research can expand the keyword list by including machine learning, neural networks, radiomics, CAD, transformers, and foundation models in their studies; conjoin WoSCC and PubMed, Scopus, and IEEE Xplore; and broaden the period of study to include the shift from traditional CAD to deep learning. The analyses may be carried out by country, institution, journal, author, or funding organization.

In terms of improving the overall research aspect, the next stage should ensure that multimodal systems undergo external validation as well as clinical assessment. There are some promising approaches, such as predictive analytics over screening cycles, cross-examination of medical images and pathology samples, quantitative examination of the tumor microenvironment, treatment outcome prediction, health equity measurement, conduct of prospective clinical trials of screening and triage techniques, and well-defined regulatory policies on LLMs and foundation models. The key challenge is to prove consistent value in routine clinical practice.

8. Conclusions

In terms of CiteSpace, research on deep learning in breast cancer revolves around a well-established imaging core and a budding frontier in precision oncology. The prime data dimension includes mammography, ultrasound, MRI, and digital pathology. Meanwhile, the application of CNNs remains a common approach in research involving deep learning. The network also shows emerging trends in the areas of response prediction, precision medicine, and clinical applications of artificial intelligence. Noteworthy clusters associated with antibody-drug conjugates as well as treatment results and applications of LLMs imply that the field is becoming more oriented towards decision-making support than mere detection. The transition implies that the validation has to be performed externally and sources have to be reproducible, equitable, and demonstrate potential benefits in terms of the outcomes of treatment.

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424. doi:10.3322/caac.21492.

[2] Giaquinto AN, Sung H, Miller KD, et al. Breast cancer statistics, 2022. CA Cancer J Clin. 2022;72:524–541. doi:10.3322/caac.21754.

[3] Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell. 2020;2:e200029. doi:10.1148/ryai.2020200029.

[4] Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet. 2019;393:1577–1579. doi:10.1016/S0140-6736(19)30037-6.

[5] Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26:1364–1374. doi:10.1038/s41591-020-1034-x.

[6] Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci Technol. 2006;57:359–377. doi:10.1002/asi.20317.

[7] Kleinberg J. Bursty and hierarchical structure in streams. Data Min Knowl Discov. 2003;7:373–397. doi:10.1023/A:1024940629314.

[8] Couture HD, Williams LA, Geradts J, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. NPJ Breast Cancer. 2018;4:30. doi:10.1038/s41523-018-0079-1.

[9] Veta M, Heng YJ, Stathonikos N, et al. Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. Med Image Anal. 2019;54:111–121. doi:10.1016/j.media.2019.02.012.

[10] Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole-slide images. Nat Med. 2019;25:1301–1309. doi:10.1038/s41591-019-0508-1.

[11] Bejnordi BE, Mullooly M, Pfeiffer RM, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. Mod Pathol. 2018;31:1502–1512. doi:10.1038/s41379-018-0073-z.

[12] Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol. 2019;20:e253–e261. doi:10.1016/S1470-2045(19)30154-8.

[13] Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. Nat Rev Clin Oncol. 2019;16:703–715. doi:10.1038/s41571-019-0252-y.

[14] Al-Antari MA, Al-Masni MA, Choi MT, Han SM, Kim TS. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int J Med Inform. 2018;117:44–54. doi:10.1016/j.ijmedinf.2018.06.003.

[15] Zhou LQ, Wu XL, Huang SY, et al. Lymph node metastasis prediction from primary breast cancer US images using deep learning. Radiology. 2020;294:19–28. doi:10.1148/radiol.2019190372.

[16] Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. J Natl Cancer Inst. 2019;111:916–922. doi:10.1093/jnci/djy222.

[17] Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. Eur Radiol. 2019;29:4825–4832. doi:10.1007/s00330-019-06186-9.

[18] Braman N, Prasanna P, Whitney J, et al. Association of peritumoral radiomics with tumor biology and pathologic response to preoperative targeted therapy for HER2-positive breast cancer. JAMA Netw Open. 2019;2:e192561. doi:10.1001/jamanetworkopen.2019.2561.

[19] Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology. 2019;292:60–66. doi:10.1148/radiol.2019182716.

[20] Lång K, Dustler M, Dahlblom V, Åkesson A, Andersson I, Zackrisson S. Identifying normal mammograms in a large screening population using artificial intelligence. Eur Radiol. 2021;31:1687–1692. doi:10.1007/s00330-020-07165-1.

[21] Larsen M, Aglen CF, Lee CI, et al. Artificial intelligence evaluation of 122,969 mammography examinations from a population-based screening program. Radiology. 2022;303:502–511. doi:10.1148/radiol.212381.

[22] Mann RM, Athanasiou A, Baltzer PAT, et al. Breast cancer screening in women with extremely dense breasts: recommendations of the European Society of Breast Imaging (EUSOBI). Eur Radiol. 2022;32:4036–4045. doi:10.1007/s00330-022-08617-6.

[23] Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. Nature. 2021;594:106–110. doi:10.1038/s41586-021-03512-4. [VERIFY: the burst figure instead lists Nat Biomed Eng 2021;5:555, doi:10.1038/s41551-020-00682-w; replace this entry with the exact WoS record.]

[24] Lamb LR, Lehman CD, Gastounioti A, Conant EF, Bahl M. Artificial intelligence (AI) for screening mammography, from the AI special series on AI applications. AJR Am J Roentgenol. 2022;219:369–380. doi:10.2214/AJR.21.27071.

[25] Mango VL, Sun M, Wynn RT, Ha R. Should we ignore, follow, or biopsy? Impact of artificial intelligence decision support on breast ultrasound lesion assessment. AJR Am J Roentgenol. 2020;214:1445–1452. doi:10.2214/AJR.19.21872.

[26] Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. Lancet. 2021;397:1750–1769. doi:10.1016/S0140-6736(20)32381-3.