

Original Research

# Landscape of Artificial Intelligence in Breast Cancer (2000–2021): A Bibliometric Analysis

Yujie Zhang<sup>1,\*</sup>, Chaoran Yu<sup>2,†</sup>, Feng Zhao<sup>2</sup>, Hua Xu<sup>2</sup>, Chenfang Zhu<sup>2,\*</sup>, Yousheng Li<sup>2,\*</sup>

<sup>1</sup>Department of Gastrointestinal Surgery, Tongji Hospital, Tongji Medical College in Huazhong University of Science and Technology, 430030 Wuhan, Hubei, China

<sup>2</sup>Department of General Surgery, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, 200240 Shanghai, China

\*Correspondence: [yujiezhong@outlook.com](mailto:yujiezhong@outlook.com) (Yujie Zhang); [guttx@hotmail.com](mailto:guttx@hotmail.com) (Yousheng Li); [sammizz1977@126.com](mailto:sammizz1977@126.com) (Chenfang Zhu)

†These authors contributed equally.

Academic Editor: Stergios Boussios

Submitted: 12 June 2022 Revised: 26 June 2022 Accepted: 28 June 2022 Published: 18 July 2022

## Abstract

**Background:** Breast cancer remains one of the leading malignancies in women with distinct clinical heterogeneity and intense multidisciplinary cooperation. Remarkable progresses have been made in artificial intelligence (AI). A bibliometric analysis was taken to characterize the current picture of development of AI in breast cancer. **Materials and Methods:** Search process was performed in the Web of Science Core Collection database with analysis and visualization performed by R software, VOSviewer, CiteSpace and Gephi. Latent Dirichlet Allocation (LDA), a machine learning based algorithm, was used for analysis of topic terms. **Results:** A total of 511 publications in the field of AI in breast cancer were retrieved between 2000 to 2021. A total of 103 publications were from USA with 2482 citations, making USA the leading country in the field of AI in breast cancer, followed by China. Mem Sloan Kettering Canc Ctr, Radboud Univ Nijmegen, Peking Univ, Sichuan Univ, ScreenPoint Med BV, Lund Univ, Duke Univ, Univ Chicago, Harvard Med Sch and Univ Texas MD Anderson Canc Ctr were the leading institutions in the field of AI in breast cancer. AI, breast cancer and classification, mammography were the leading keywords. LDA topic modeling identified top fifty topics relating the AI in breast cancer. A total of five primary clusters were found within the network of fifty topics, including radiology feature, lymph node diagnosis and model, pathological tissue and image, dataset classification and machine learning, gene expression and survival. **Conclusions:** This research depicted AI studies in breast cancer and presented insightful topic terms with future perspective.

**Keywords:** breast cancer; bibliometric analysis; artificial intelligence; Web of Science

## 1. Introduction

Breast cancer remains one of the leading malignancies in women featuring by distinct clinical heterogeneity and multidisciplinary management with second leading cause of death [1–3]. Remarkable progresses in several aspects have been made in this field, such as image-based system screening, genomic characterization and artificial intelligence (AI) [1,4,5]. In fact, advances in computers and imaging systems enable the rising use of AI in clinical diagnosis, treatment and prediction of breast cancer. The ability of AI to deal with increasing volume of data has significantly outweighed human efforts. Specifically, AI offers a well-built strategy for accurate model demarcation, characterization extractions, clinical phenotypes and risks classifications [6,7].

Over the past twenty years, techniques of computer-aided detections and diagnosis have been revolutionized by AI, such as automated breast cancer detection of mammography and breast tomosynthesis, particularly driven by machine learning and deep learning algorithms [8]. Nations-across digital mammography examinations shows that the performance of AI is statistically noninferior to the average of 101 radiology experts [9]. Meanwhile, radiologists with

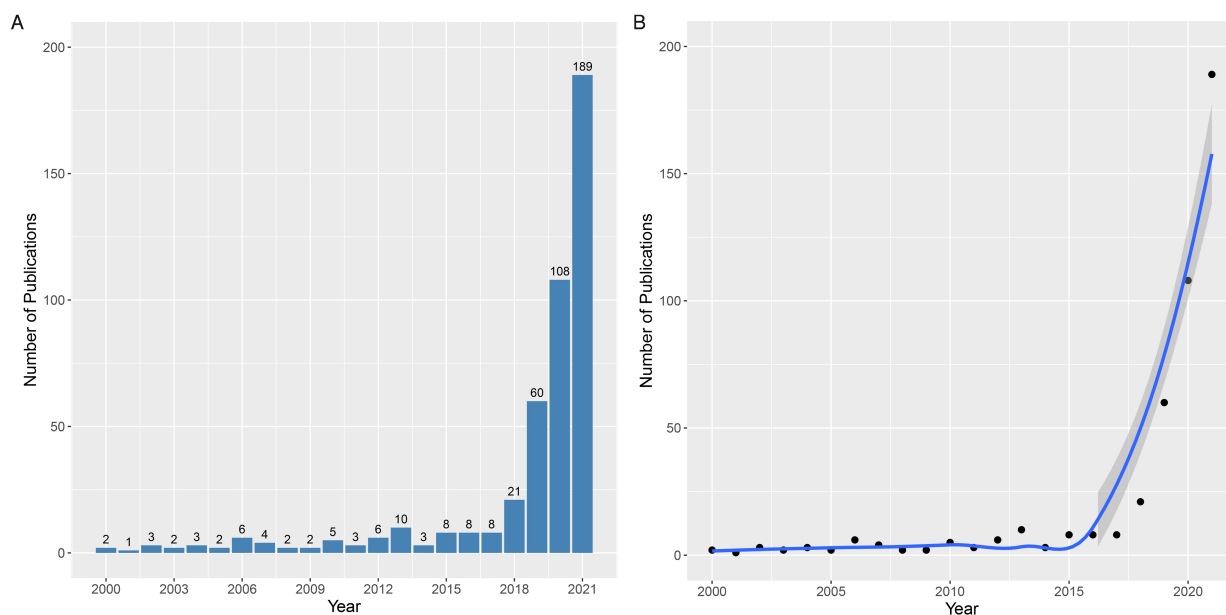
AI support display significant improvement of sensitivity and specificity comparing to those without AI support [10]. Advances of AI also accelerate the improvement of digital histopathological diagnosis, with more accurate detection, classification and prediction [11].

Increasing interests of bibliometric analysis have been noticed in several aspects, such as periodontology, gynecology and neurology [12–14]. However, similar study in AI of breast cancer remains sparse. Hereby, based on mounting studies from 2000 to 2021 in this field, a bibliometric analysis was taken to fully characterize the current picture of development of AI in breast cancer, highlight potential key topics via machine learning based Latent Dirichlet allocation (LDA) model algorithms and provide insightful association between AI and breast cancer.

## 2. Materials and Methods

Search process was performed in the Web of Science Core Collection database (<https://clarivate.com/webofsciencegroup/solutions/web-of-science-core-collection/>) [15]. Inclusion criteria was (1) studies with listed topic words ((breast) AND (cancer OR tumor OR carcinoma OR neoplasm) AND (artificial intelligence)); (2) publication





**Fig. 1. Annual number of publications of artificial intelligence (AI) in breast cancer and a model fitting curve.** (A) Number of publications in each year from 2000 to 2021. (B) A model curve fitting the growth of the publications.

year was between 2000 to 2021; (3) language of studies was English; (4) only articles were included for analysis.

R software (version 4.1.1, Auckland University, New Zealand, <https://www.r-project.org/>) (bibliometrix package), VOSviewer (version 1.6.18, Leiden University, the Netherlands, <https://www.vosviewer.com/>) and CiteSpace (version 5.8.R3, Drexel University, Philadelphia, PA, USA, <https://citespace.podia.com/>) were utilized for data processing and visualization [16–21]. A machine learning based topic modeling algorithm, Latent Dirichlet Allocation (LDA), was used to train and determine fifty most essential topics highly related to the publications of AI in breast cancer and the results were visualized by Gephi Software (version 0.9.5, Compiègne, France, <https://gephi.org>).

### 3. Results

511 publications relating to AI in breast cancer were retrieved from the WoS database between 2000 to 2021 (Fig. 1A,B).

Distinct country-specific publications were displayed ranking from top. A total of 103 publications were from USA with 2482 citations, making USA the leading country in the field of AI in breast cancer (Fig. 2A–C).

China published 79 studies as the second most country, but with the third most citation. To further characterize the publications patterns, a dual-map thematic overlays was used for portfolio analysis. The results showed that medicine, medical clinical fields and molecular, biology, genetics fields were two highly correlated categories with intense publication interests and citation focus (Fig. 3).

Other fields, such as mathematics, systems, mathematical, may not be under fully development, although AI

was indeed originated from such field.

The network of contributing countries and global institutions were demonstrated (Fig. 4A,B).

In fact, Mem Sloan Kettering Canc Ctr, Radboud Univ Nijmegen, Peking Univ, Sichuan Univ, ScreenPoint Med BV, Lund Univ, Duke Univ, Univ Chicago, Harvard Med Sch and Univ Texas MD Anderson Canc Ctr were the leading institutions in the field of AI in breast cancer.

Top ten most influential journals were identified (Table 1).

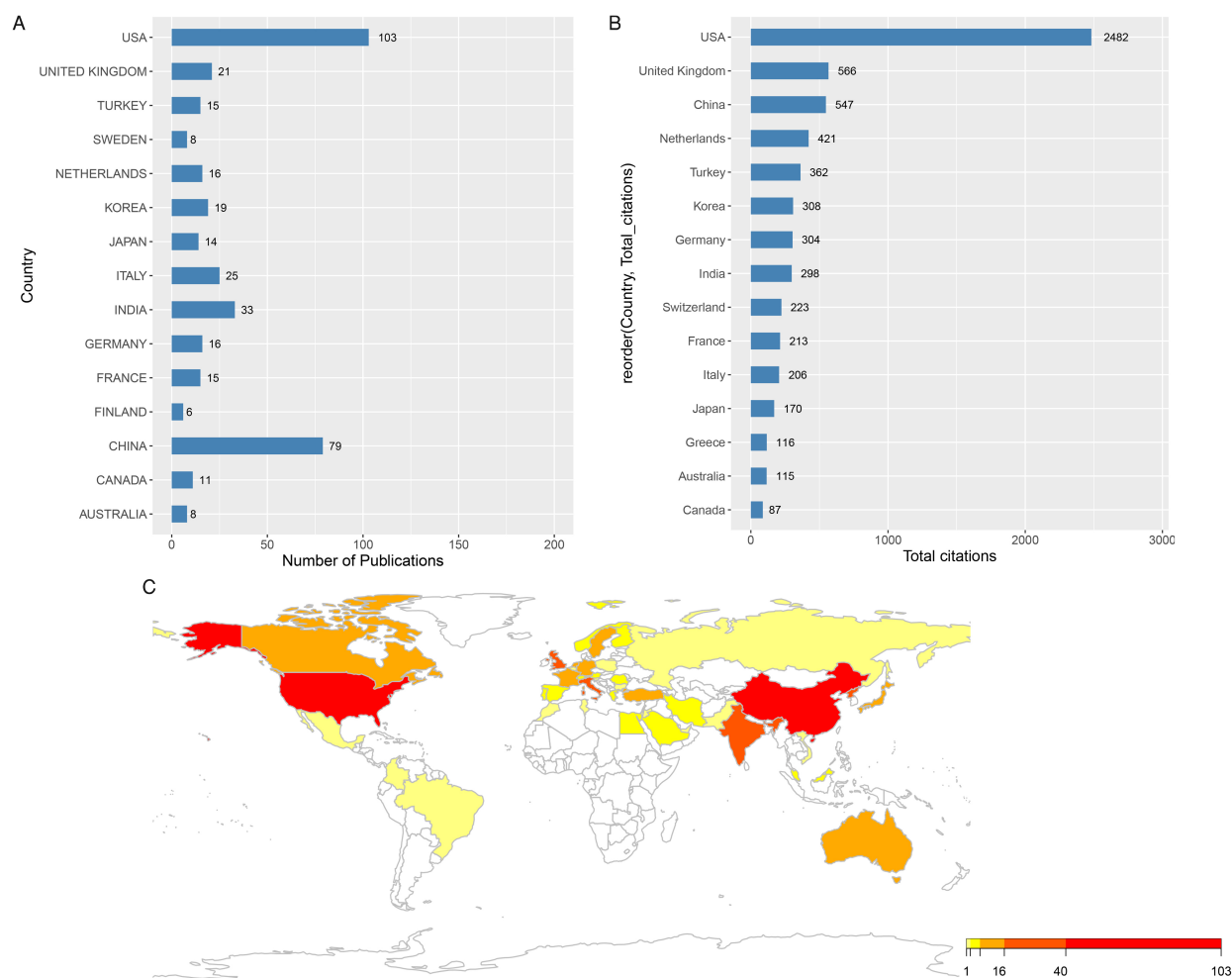
To further delineate the keywords features, a co-occurrence network of the most frequent keywords were extracted and established (Fig. 5).

AI, breast cancer and classification, mammography were the leading keywords. However, such demonstration may not help identify potential significant keywords topics. Therefore, a LDA topic modeling was used to generate the top fifty topics relating the AI in breast cancer (Fig. 6).

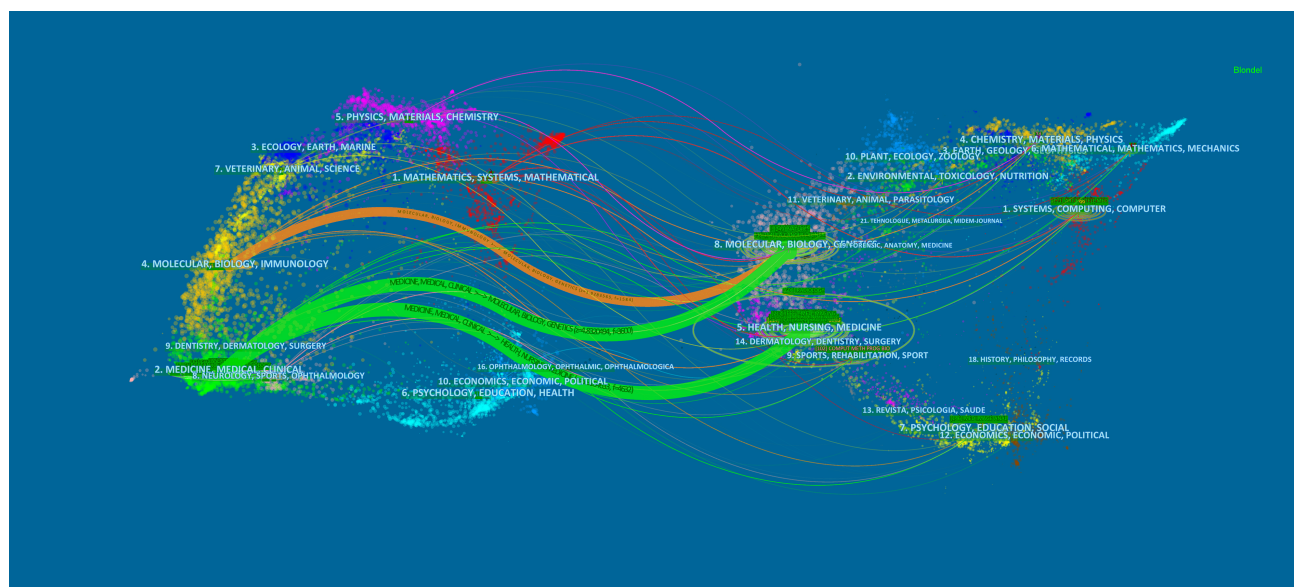
The inter-correlation of those topics was also revealed by a clustering network. A total of five primary clusters were found, including radiology feature, lymph node diagnosis and model, pathological tissue and image, dataset classification and machine learning, gene expression and survival (Fig. 7).

### 4. Discussion

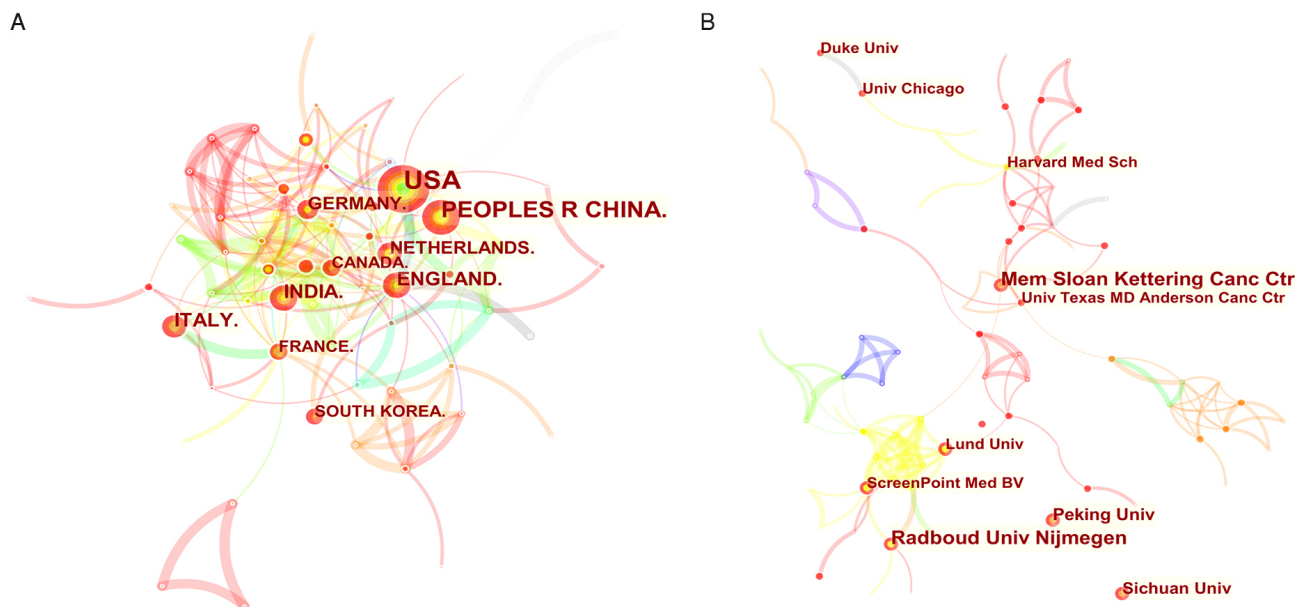
This study reported that, with 511 publications relating to AI in breast cancer and machine learning techniques, five primary cluster of distinct bibliometric topics were identified, including radiology feature, lymph node diagnosis and model, pathological tissue and image, dataset classification and machine learning, gene expression and survival.



**Fig. 2. Total publications and citations of each country of AI in breast cancer.** (A) Total number of publications in each country from 2000 to 2021. (B) Total citations of publications in each country from 2000 to 2021. (C) The number of publications of each country shown in a world map.



**Fig. 3. Publication portfolio of AI in breast cancer.**



**Fig. 4. Contributing network of countries and global institutions.** (A) Network of contributing countries in the field of AI in breast cancer, with top countries including USA, China, England, India, Germany, Netherlands, Italy, Canada, France and South Korea. (B) Network of contributing global institutions in this field, including Mem Sloan Kettering Canc Ctr, Radboud Univ Nijmegen, Peking Univ, Sichuan Univ, ScreenPoint Med BV, Lund Univ, Duke Univ, Univ Chicago, Harvard Med Sch and Univ Texas MD Anderson Canc Ctr.

**Table 1. The top 10 journals relating to AI in breast cancer.**

Rank	Journal	Number (n)	Country	Citations	Average citation	2020 IF*	Quartile in category
1	Breast	15	England	234	15.6	4.38	Q1/Q2
2	European Radiology	13	USA	141	10.846	5.315	Q1
3	Scientific Reports	10	England	123	12.3	4.38	Q1
4	Cancers	8	Switzerland	64	8	6.639	Q1
5	Applied Sciences-basel	7	Switzerland	27	3.857	2.679	Q2/Q3
6	Diagnostics	7	Switzerland	24	3.429	3.706	Q2
7	European Journal of Radiology	7	Ireland	19	2.714	3.528	Q2
8	Frontiers in Oncology	7	Switzerland	69	9.857	6.244	Q2
9	Ieee Access	7	USA	31	4.429	3.367	Q2
10	Journal of Digital Imaging	7	USA	80	11.429	4.056	Q1

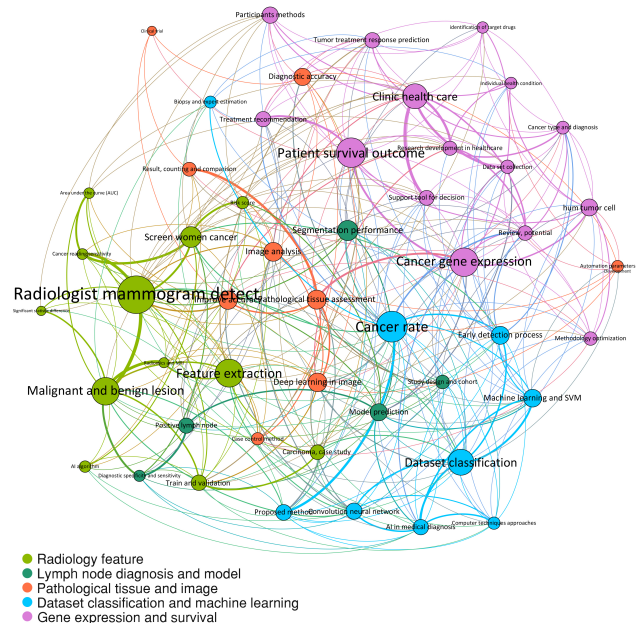
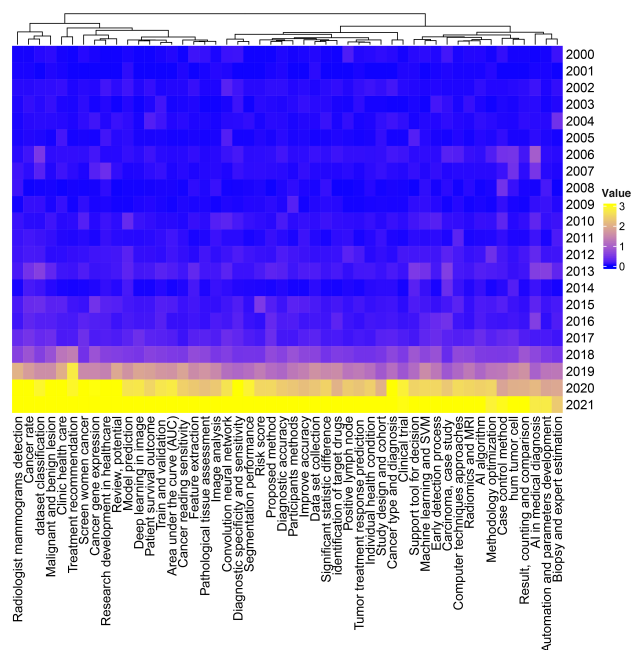
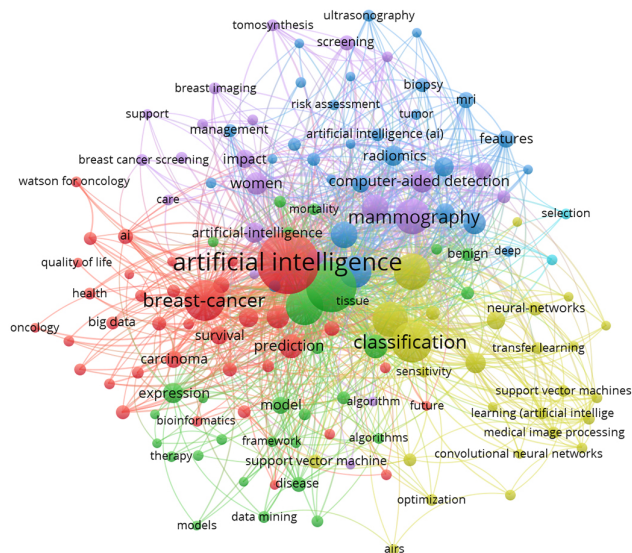
\*The impact factors (IF) of journals were obtained from the 2020 Web of Science Journal Citation Reports (JCR).

Radiology features covers some of the most intense research interests, such as radiologist mammogram detect, feature extraction and AI algorithm. Of note, introduction of computer supporting system to mammography was taken place back in 1990s [22,23]. However, the real uprising value of this field is largely contributed by the success of deep learning [9,10,24]. Alejandro Rodriguez-Ruiz *et al.* [9] established an AI system for digital mammography evaluation with a comparable accuracy. The system was built based on deep learning convolutional neural networks and image-based feature classifiers with over 9000 training malignancies and 180 000 normal cases. Similar AI algorithm, deep neural networks were also employed in other study [24]. With over 1 000 000 images, Wu *et al.* [24] presented a network for breast cancer screening exam classification with a well validated performance as the area under

the curve (AUC) reaching to 0.895. The model architecture was trained and built by four variant models and generated two primary modules, one for view-specific columns, the other for mapping the hidden representations via connected layers.

AI in medical diagnosis for early detection process using algorithms such as convolution neural network and other machine learning methods demonstrates a promising attractive topic. Convolution neural network was among the primary cluster of dataset classification and machine learning, it was indeed one of the AI algorithms most commonly used in the field of AI in breast cancer. It was inspired by biological visual cortex with remarkable breakthrough made in the last decades [25,26]. The prevailing technique of multi-view deep convolutional neural networks commonly used in breast cancer were initially de-





**Fig. 7. Topics network with clustering.** All the determined topics were clustered and colored by Gephi software. Five clusters were identified, including radiology feature (light green), lymph node diagnosis and model (dark green), pathological tissue and image (orange), dataset classification and machine learning (light blue), gene expression and survival (violet).

convolutional layers with concatenated vectors, the other to formulate a fully connected layer for output distribution. It targeted high-resolution medical images without heavily downscaling image information extraction, thereby making it adequate to upgrade network architecture with accurate prediction.

Machine learning based LDA model is one of the mostly used context analytic methods with easy accessibility and feasibility. By text-mining of large number of publications, LDA presents most significantly enriched topic terms and contents worthy further discussion. Similar attempts have been made in several fields [29–31].

Limitations remain, most of studies rely on large volume of digitalized data, but with comparable less sample size. Moreover, actual clinical practice, such as physical examinations, laboratory data and patient-physical communications remain mostly untouched in this field. Multi-dimensional data resources and environment enable stronger evidence of AI prediction. Meanwhile, long term follow-up of clinical outcomes prediction in breast cancer is also yet to be fully investigated. Therapeutic management, identification of novel drugs, clinical trial navigators, clinical practice assistance as well as automated individual-specific clinical healthcare systems are potential clinical applications in the field of AI in breast cancer.

This study underscores the advantages of machine learning based algorithms for topic modeling in the field of AI in breast cancer. The research categories of AI in

breast cancer characterized by this study are not only supported by increasing research publications but also beneficial to opportunity of better alignment to future direction and specific needs. This study suggests at least five main areas to be highlighted by future investigations, including (1) radiology feature, (2) lymph node diagnosis and model, (3) pathological tissue and image, (4) dataset classification and machine learning, (5) gene expression and survival. Those focuses create future collaborative research agenda as well as identification of funding organizations. In fact, equity of funding distribution remains a key issue among research agencies.

## 5. Conclusions

This research depicted AI studies in breast cancer and presented insightful topic terms with future perspective.

## Author Contributions

YZ, CY and CZ carried out data analysis. CY, FZ, HX and YL drafted the manuscript. All authors participated in study design and data collection. All authors read and approved the final manuscript.

## Ethics Approval and Consent to Participate

Not applicable.

## Acknowledgment

Not applicable.

## Funding

This research received no external funding.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, *et al.* International evaluation of an AI system for breast cancer screening. *Nature*. 2020; 577: 89–94.
- [2] Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, *et al.* Breast cancer. *Nature Reviews Disease Primers*. 2019; 5: 66.
- [3] 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: A Cancer Journal for Clinicians*. 2018; 68: 394–424.
- [4] Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DSM, Kerkliowske K, *et al.* National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology*. 2017; 283: 49–58.
- [5] Liang Y, Zhang H, Song X, Yang Q. Metastatic heterogeneity of breast cancer: Molecular mechanism and potential therapeutic targets. *Seminars in Cancer Biology*. 2020; 60: 14–27.
- [6] Sheth D, Giger ML. Artificial intelligence in the interpretation of breast cancer on MRI. *Journal of Magnetic Resonance Imaging*. 2020; 51: 1310–1324.
- [7] Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. *Clinical Radiology*. 2019; 74: 357–366.
- [8] Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Seminars in Cancer Biology*. 2021; 72: 214–225.
- [9] Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, *et al.* Stand-alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison with 101 Radiologists. *Journal of the National Cancer Institute*. 2019; 111: 916–922.
- [10] Rodríguez-Ruiz A, Krupinski E, Mordang J, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, *et al.* Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. *Radiology*. 2019; 290: 305–314.
- [11] Ibrahim A, Gamble P, Jaroensri R, Abdelsamea MM, Mermel CH, Chen PC, *et al.* Artificial intelligence in digital breast pathology: Techniques and applications. *The Breast*. 2020; 49: 267–273.
- [12] Ahmad P, Slots J. A bibliometric analysis of periodontology. *Periodontology 2000*. 2021; 85: 237–240.
- [13] Brandt JS, Hadaya O, Schuster M, Rosen T, Sauer MV, Ananth CV. A Bibliometric Analysis of top-Cited Journal Articles in Obstetrics and Gynecology. *JAMA Network Open*. 2019; 2: e1918007.
- [14] Ramos MB, Koterba E, Rosi Júnior J, Teixeira MJ, Figueiredo EG. A Bibliometric Analysis of the most Cited Articles in Neurocritical Care Research. *Neurocritical Care*. 2019; 31: 365–372.
- [15] Clarivate A. Web of science. Clarivate Analytics, 2019.
- [16] Aria M, Cuccurullo C. Bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics*. 2017; 11: 959–975.
- [17] Campra M, Riva P, Oricchio G, Brescia V. Bibliometrix analysis of medical tourism. *Health Services Management Research*. 2021; 9514848211011738. (in press)
- [18] Perianes-Rodriguez A, Waltman L, van Eck NJ. Constructing bibliometric networks: a comparison between full and fractional counting. *Journal of Informetrics*. 2016; 10: 1178–1195.
- [19] Chen C. A Glimpse of the First Eight Months of the COVID-19 Literature on Microsoft Academic Graph: Themes, Citation Contexts, and Uncertainties. *Frontiers in Research Metrics and Analytics*. 2020; 5: 607286.
- [20] Chen C, Song M. Visualizing a field of research: A methodology of systematic scientometric reviews. *PLoS ONE*. 2019; 14: e0223994.
- [21] Chen C, Leydesdorff L. Patterns of connections and movements in dual-map overlays: a new method of publication portfolio analysis. *Journal of the Association for Information Science and Technology*. 2014; 65: 334–351.
- [22] Zheng B, Chang Y, Staiger M, Good W, Gur D. Computer-aided detection of clustered microcalcifications in digitized mammograms. *Academic Radiology*. 1995; 2: 655–662.
- [23] Chan H, Doi K, Vybrony CJ, Schmidt RA, Metz CE, Lam KL, *et al.* Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis. *Investigative Radiology*. 1990; 25: 1102–1110.
- [24] Wu N, Phang J, Park J, Shen Y, Huang Z, Zorin M, *et al.* Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. *IEEE Transactions on Medical Imaging*. 2020; 39: 1184–1194.
- [25] Lawrence S, Giles CL, Ah Chung Tsoi, Back AD. Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks*. 1997; 8: 98–113.
- [26] Vedaldi A, Lenc K. MatConvNet: Convolutional neural networks for matlab. The 23<sup>rd</sup> ACM International Conference on Multimedia (pp. 689–692). ACM. 2015.

- [27] Geras KJ, Wolfson S, Shen Y, Wu N, Kim S, Kim E, *et al.* High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv preprint arXiv:1703.07047. 2017.
- [28] Rawat W, Wang Z. Deep Convolutional Neural Networks for Image Classification: a Comprehensive Review. *Neural Computation*. 2017; 29: 2352–2449.
- [29] Tran BX, Latkin CA, Sharafeldin N, Nguyen K, Vu GT, Tam WWS, *et al.* Characterizing artificial intelligence applications in cancer research: a latent dirichlet allocation analysis. *JMIR Medical Informatics*. 2019; 7: e14401.
- [30] Escobar KM, Vicente-Villardón JL, de la Hoz-M J, Useche-Castro LM, Alarcón Cano DF, Siteneski A, *et al.* Frequency of Neuroendocrine Tumor Studies: Using Latent Dirichlet allocation and HJ-Biplot Statistical Methods. *Mathematics*. 2021; 9: 2281.
- [31] Matsutani T, Ueno Y, Fukunaga T, Hamada M. Discovering novel mutation signatures by latent Dirichlet allocation with variational Bayes inference. *Bioinformatics*. 2019; 35: 4543–4552.