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

Yujie Zhang1,*,†, Chaoran Yu1,†, Feng Zhao2, Hua Xu2, Chenfang Zhu2,*, Yousheng Li2,*

1Department of Gastrointestinal Surgery, Tongji Hospital, Tongji Medical College in Huazhong University of Science and Technology, 430030 Wuhan, Hubei, China
2Department of General Surgery, Shanghai Ninth People’s Hospital, Shanghai Jiao Tong University School of Medicine, 200240 Shanghai, China
*Correspondence: yujiezhang@outlook.com (Yujie Zhang); gutx@hotmail.com (Yousheng Li); sammizzi1977@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/webofangroup/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...
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.

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

*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. 5. Co-occurrence network map of the most frequent 145 keywords. All the keywords were visualized and color-clustered.

Fig. 6. Distribution value of fifty topics determined by Latent Dirichlet Allocation (LDA) in the field of AI in breast cancer from 2000 to 2021. Fifty topics were processed and analyzed by topic-modeling machine-learning based algorithm. All the topics were visualized in heat map. Blue: low value; yellow: high value.

signed by Geras et al. [27] in 2017. A deep convolution neural network reclassifies an input image into multiple colors channels and processes the pixel-level image with nonlinear functions, then outputs the probability distribution via multi-layer perceptron [28]. The multi-view deep convolution neural network proposed by Geras et al. [27] consisted of two steps, one to perform view-specific 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


