Top 100 most-cited publications on breast cancer and machine learning research

Hanis, Tengku Muhammad; Islam, Md Asiful; Musa, Kamarul Imran

DOI:
10.2174/0929867328666211108110731

Document Version
Peer reviewed version

Citation for published version (Harvard):
Hanis, TM, Islam, MA & Musa, KI 2022, 'Top 100 most-cited publications on breast cancer and machine learning research: a bibliometric analysis', *Current medicinal chemistry*, vol. 20, no. 8, pp. 1426-1435. https://doi.org/10.2174/0929867328666211108110731

Send Orders for Reprints to reprints@benthamscience.ae

Current Medicinal Chemistry, 2019, *Volume, Page Enation*

ARTICLE TYPE

Title: The Top 100 Most-Cited Publications on Breast Cancer and Machine Learning Research: A Bibliometric Analysis

Tengku Muhammad Hanis[a], Md Asiful Islam[b*], Kamarul Imran Musa[a*]

[a]*Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Malaysia;*
[b]*Department of Haematology, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Malaysia*

ARTICLE HISTORY

Received:
Revised:
Accepted:

DOI:

Abstract:

Background: Rapid advancement in computing technology and digital information lead to the possible use of machine learning on breast cancer.

Objective: This study aims to evaluate the research output of the top 100 publications and further identify a research theme of breast cancer and machine learning studies.

Methods: The two databases; Scopus and Web of Science, were used to extract the top 100 publications. The top 100 publications were filtered based on the total citation of each paper. Additionally, a bibliometric analysis was applied to the top 100 publications.

Results: The top 100 publications were ranged between 1993 and 2019. The most productive author was Giger ML, and the top two institutions were the University of Chicago and the National University of Singapore. The most active countries were the USA, Germany, and China. There were ten clusters identified as both basic and specialised themes of breast cancer and machine learning.

Conclusion: There was an equal interest from countries of various regions in breast cancer and machine learning research. In fact, a few Asian countries, such as China, India, and Singapore, were listed in the top ten countries based on the total citation. Additionally, the use of deep learning and breast imaging data was trending in the past ten years of breast cancer and machine learning research.

Keywords: bibliometrics, breast cancer, machine learning, research trend, research output, research productivity.

1. INTRODUCTION

Breast cancer has been the most common cancer among women affecting 25% of women worldwide [1]. The incidence of breast cancer is higher in developed countries compared to developing countries [2]. However, the difference is most likely to be overemphasised, as there is a lack of screening in developing countries [3]. Additionally, the incidence of breast cancer was reported to steadily increase over the years in South America, Africa, and Asia [2]. A study published in 2019 estimated the incidence of breast cancer to be increased by 23.1% in 2020 and 60.7% by 2025 [4].

* Corresponding authors:
E-mails: asiful@usm.my or ayoncx70@yahoo.com (MAI) and drkamarul@usm.my (KIM).

Machine learning is a branch of artificial intelligence (AI). It is one of the most thriving subfield areas of AI due to an advance in computer electronic and digital information. The main aim of machine learning is to make predictions and classifications based on a pattern learned from a dataset. Subsequently, this prediction and classification are used to support the decision-making process. Since 2006, deep learning or representative learning had gained popularity and has been used in many areas, including medicine and healthcare [5]. Deep learning is a subfield area of machine learning which enables a machine to learn a representation from raw data and provides a better prediction and classification [6].

Advancement in technology enables efficient data storing and processing. This abundance of data makes it possible for the use of machine learning in medicine and healthcare. The uses of machine learning in medical research includes a prediction of Alzheimer's disease [7], prediction of kidney transplant survival [8], prediction of late-onset preeclampsia [9], prediction of pulmonary hypertension [10], detection of a skin lesion in the diagnosis of melanoma [11], detection of subclinical depressive symptoms among healthy individuals

0929-8673/19 $58.00+.00

[12], and prediction of kidney injury after liver cancer resection [13]. Additionally, there were various types of breast cancer research conducted, and the possible use of machine learning in the area is endless. A common application of machine learning in breast cancer primarily involves cancer prediction and prognosis using various datasets such as medical history, clinical data, imaging data, histopathological data and genomic data. However, machine learning had yet to be fully implemented in medical practice, though, physicians and healthcare professionals are aware of the benefit of its implementation [14].

Bibliometrics is a study of a collection of literature using a quantitative approach. Bibliometrics has been used since the 1920s [15]. However, the surge of publications each year makes the bibliometric analysis more relevant than ever. The bibliometric analysis provides a good measure to assess the advance of research in any area of study. A citation is sourced from a reference in a publication, and it reflects the scientific significance of a paper [16]. Also, a citation may reflect a trend and theme of an area of research as most authors in that particular research area cite a similar collection of publications. However, a citation reflects neither the quality of a study nor its impact on policy and decision making.

This study aims to assess the research output in breast cancer and machine learning in terms of the distribution of publications across countries, institutions, journals, and authors related to breast cancer and machine learning area. Secondly, this study aims to determine the themes and trends in breast cancer and machine learning research.

2. MATERIALS AND METHOD

2.1. Study search and selection

All scientific publications related to breast cancer and machine learning were retrieved from Scopus and Web of Science databases on 1st November 2020. The search terms are included in Supplement Table 1. Both results were limited to the research article, proceeding, and review. Metadata, titles, and abstracts were downloaded for each publication in a BibTex file format. Both datasets were combined and sorted based on the total citation. Duplicates with a lower citation were removed. The publications were manually screened, starting with the highest cited publication until a top hundred most cited publications were reached. During the screening process, studies related to only breast cancer and machine learning area were included.

2.2. Data analysis

Data management and analyses were done in R software version 4.0.3 [17]. Bibliometric analyses were done using Bibliometrix package version 3.0.3 [18]. Bradford's law of scattering describes the distribution of the publication in a collection of documents [19]. The publications were divided into three zones with a relatively similar number of publications in each zone. The core zone represents the core journals in this research area.

A thematic map is a version of a strategic theme proposed by Cobo *et al.* in 2011 [20] implemented in the Bibliometrix package. The map was constructed based on the words of publications' titles which frequent at five times in the dataset. The thematic map is useful in identifying research clusters in the study. The Three-fields plot is a Sankey diagram used to visualise the relationship between the top items of the three selected fields. The Three-fields plot was constructed through the Bibliometrix package.

3. RESULTS

There were 14347 publications retrieved from the Scopus database, and 10920 publications retrieved from the Web of Science database. Fig 1 presents the flow of study selection in this study.

3.1. Main characteristics of the included studies

The top 100 publications related to breast cancer and machine learning consisted of 94 research articles, four conference papers and two review papers, among which 51 of the top 100 studies were funded. The number of citations varied between 106 and 556 with an average citation per document being 151.9. The papers ranged from 1993 until 2019. The highest number of papers were published in 2003, followed by 2016 and 2018 with 13, 8, and 7 papers respectively. Fig 2 presents the distribution of the top 100 publications according to the year of publication. The most cited publications were by Bejnordi *et al.* [21], Schmidt-Kittler *et al.* [22], and Cruz and Wishart [23] with a total citation of 556, 466, and 431 respectively. Table 1 shows the top ten of the most cited publications on breast cancer and machine learning.

3.2. Distribution of authors

There were three single-authored papers, while the remaining were multi-authored papers. The number of authors for the top 100 most cited papers in breast cancer and machine learning was 1329. The number of authors per paper was ranged between 1 and 748. About 80% of the papers had ten authors or less. Additionally, the most productive authors were Giger ML with six papers, followed by Feldman M and Madabhushi M, and both published four papers each. The remaining authors had only three papers or less. Fig 3 presents the top ten most productive authors in the area of breast cancer and machine learning.

3.3. Distribution of countries

There were 34 countries involved in the publication of the top 100 most cited papers related to breast cancer and machine learning. Countries with the highest total citation were the USA with 4629 total citations, followed by Germany and China with 1546 and 1164 total citations, respectively. The top ten countries based on the total citations are presented in Table 2. Fig 4 describes the relationship between top authors, top journals, and top countries. Most publications' affiliations of the top journals were affiliated with the USA.

3.4. Distribution of institutions

There were 247 institutions affiliated with the publications in this study. The top institution based on the corresponding authors' affiliation was the University of Chicago (n=4), followed by the National University of Singapore (n=3), and the remaining institutions had two papers and less. Additionally, the University of Chicago and the National University of Singapore had been the top two institutions based on the fraction of co-authors per paper in this study area. The top ten institutions based on the fraction of co-authors per

paper in this study are presented in Table 3. Most of the top ten institutions were from the USA.

3.5. Distribution of journals

According to Bradford's law of scattering, the core journal consisted of seven journals with 33 publications. The core journals were IEEE Transactions on Medical Imaging (8 publications), Expert Systems with Applications (6 publications), Artificial Intelligence in Medicine (4 publications), Medical Physics (4 publications), Radiology (4 publications), Scientific Reports (4 publications), and Computers in Biology and Medicine (3 publications).

3.6. Collaborations

There were 1329 authors for the top 100 most cited papers. The average co-author per document was 14 with a collaboration index of 13.7. Fig 5 represents the collaborations between countries in this study area. The USA was the most active country, followed by China, whereas Canada and Germany were in third place.

3.7. Publication trends

Fig 6 depicts a thematic map of the top 100 most cited publications in this study. An upper right quadrant represents a motor theme, while a lower right quadrant represents a basic theme. An upper left quadrant reflects an isolated and highly developed theme, while a lower left quadrant reflects an emerging or declining theme. There were ten clusters identified in this study. Each cluster was represented by the top three keywords derived from the titles of publications of each corresponding cluster. There were five clusters related to the isolated and highly developed theme and another five clusters related to a basic theme (Fig 6). Fig 7 illustrates the keywords associated with each publication's database for the past ten years. The most frequent keyword was breast tumour in 2012. The most frequent keyword associated with machine learning technique was cluster analysis in 2011.

4. DISCUSSION

Most of the publications in this study were multi-authored from different disciplines, which reflected the multidisciplinary nature of breast cancer and machine learning research. The publications in this study spanned between 1993 and 2019, reflecting a strong interest in the application of machine learning on breast cancer in the last three decades. The two publications in 1993 [24, 25] involved the application of machine learning in the risk assessment of the psychosocial aspect of breast cancer patients and mammograms classification. The most recent publication in 2019 was related to the development of a decision support system for breast cancer metastases [25].

The USA was the most active country as most publications from the top journals in this study were affiliated with this country, as shown in Fig 5. Subsequently, the USA lead other countries in terms of the total number of citations. About half of the top ten countries with the highest citations were European countries. However, there were three Asian countries, namely; China, India, and Singapore, listed in the top ten countries with the highest total citations (Table 2). Additionally, Fig 5 shows the participation of researchers from the African region (Nigeria and Egypt) in this area of research. Thus, the participation of countries from different regions represents an equal interest in the use of machine learning on breast cancer.

The basic theme of breast cancer and machine learning research were shown in the lower right quadrant of Fig 6. The five clusters identified in this quadrant were the most prevalent research themes and the main research domains in this area. The most common techniques of machine learning were neural network and deep learning, which represented two clusters in this quadrant. Both techniques were related as deep learning or deep neural network is considered as the more advanced and recent technique of neural network. Deep learning was predominantly used in research related to medical image analysis. The most frequently used medical image in breast cancer research was digital mammogram [27], though MRI was predominantly used in medical diagnosis research [26]. Additionally, genomic research was another cluster of main research domains related to breast cancer and machine learning. A high degree of centrality and mid-degree density of this cluster indicated that it was recently migrated from the quadrant of a motor theme. A research cluster represented in a motor theme (upper right quadrant of Fig 6) was a most developed and active research area, though none were identified in this study. There were another two clusters identified by the thematic map in this quadrant, which were breast cancer diagnosis and analysis of breast cancer features. However, both clusters were quite equivocal, and did not give much insight into machine learning and breast cancer research. Nonetheless, the research clusters in this quadrant were studied for a long time and transversal across publications in this study.

The other five clusters in the upper left quadrant of Fig 6 were isolated and highly developed themes. Clusters in this quadrant represented a niche research area related to machine learning and breast cancer. Keywords such as imaging, genetics, and cells reflected a type of data used in this specialised area. Machine learning techniques such as support vector machines and fuzzy logic were commonly used in this specialised area of breast cancer and machine learning. Also, most studies in this area seem to be related to the assessment of tumour immunity and metastases.

In recent years, machine learning methods such as deep learning, support vector machine, and cluster analysis were famously used across the publications, as indicated in Fig 7. There were three publications [21, 28, 29] from the top ten most-cited papers that used deep learning in their studies (Table 1). The use of imaging data, such as mammograms and echocardiograms were quite frequent over the last few years. Keywords such as genomics and gene expression regulation reflected the popularity of genomic studies in breast cancer and machine learning area. Additionally, machine learning had been used in the area of histopathology of breast cancer, which could be seen by the prevalence of keywords such as cell nucleus, pathology, signal transduction and basal cell.

The main strength of this study was the use of a data-driven approach to assessing the research output, theme and trend of breast cancer and machine learning publications. Secondly, this study used the two central databases (Scopus and Web of Science) to extract the top 100 most-cited publications related to breast cancer and machine learning. However, these two databases had a different structure of

metadata which limited further analysis such as citation and co-citation analysis. Also, this study only included publication in English as it was impossible to do the analysis, including a thematic map and trending keywords with different languages included.

CONCLUSION

This study provided a bibliometric analysis of top publications on breast cancer and machine learning. The most active country was the USA, with the rise of Asian countries such as China, India, and Singapore. The most popular method was deep learning, and the most used data were imaging data.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The publication of this study was supported by the School of Medical Sciences, Universiti Sains Malaysia.

SUPPORTIVE/SUPPLEMENTARY MATERIAL

Supplement Table 1. Search terms used for Scopus and Web of Science databases.

Supplement Table 2. Top 100 research articles included in the study.

REFERENCES

[1] Parks, R. M.; Derks, M. G. M.; Bastiaannet, E.; Cheung, K. L. Breast Cancer Epidemiology. In *Breast Cancer Management for Surgeons*; Springer, Cham, 2018; pp 5615–5623. https://doi.org/10.1007/978-3-319-56673-3.

[2] Goncheh, M.; Pournamdar, Z.; Salehiniya, H. Incidence and Mortality and Epidemiology of Breast Cancer in the World. *Asian Pacific J. Cancer Prev.* **2016**, *17* (S3), 43–46. https://doi.org/10.7314/APJCP.2016.17.S3.43.

[3] Rojas, K.; Stuckey, A. Breast Cancer Epidemiology and Risk Factors. *Clin. Obstet. Gynecol.* **2016**, *59* (4), 651–672. https://doi.org/10.1097/GRF.0000000000000239.

[4] Zaheer, S.; Shah, N.; Maqbool, S. A.; Soomro, N. M. Estimates of Past and Future Time Trends in Age-Specific Breast Cancer Incidence among Women in Karachi, Pakistan: 2004-2025. *BMC Public Health* **2019**, *19* (1), 1–9. https://doi.org/10.1186/s12889-019-7330-z.

[5] Al-Jarrah, O. Y.; Yoo, P. D.; Muhaidat, S.; Karagiannidis, G. K.; Taha, K. Efficient Machine Learning for Big Data: A Review. *Big Data Res.* **2015**, *2* (3), 87–93. https://doi.org/10.1016/j.bdr.2015.04.001.

[6] Lecun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521* (7553), 436–444. https://doi.org/10.1038/nature14539.

[7] Zhang, Y.; Wang, S. Detection of Alzheimer's Disease by Displacement Field and Machine Learning. *PeerJ* **2015**, *2015* (9), 1–29. https://doi.org/10.7717/peerj.1251.

[8] Mark, E.; Goldsman, D.; Gurbaxani, B.; Keskinocak, P.; Sokol, J. Using Machine Learning and an Ensemble of Methods to Predict Kidney Transplant Survival. *PLoS One* **2019**, *14* (1), 1–13. https://doi.org/10.1371/journal.pone.0209068.

[9] Jhee, J. H.; Lee, S.; Park, Y.; Lee, S. E.; Kim, Y. A.; Kang, S. W.; Kwon, J. Y.; Park, J. T. Prediction Model Development of Late-Onset Preeclampsia Using Machine Learning-Based Methods. *PLoS One* **2019**, *14* (8), 1–12. https://doi.org/10.1371/journal.pone.0221202.

[10] Leha, A.; Hellenkamp, K.; Unsöld, B.; Mushemi-Blake, S.; Shah, A. M.; Hasenfuß, G.; Seidler, T. A Machine Learning Approach for the Prediction of Pulmonary Hypertension. *PLoS One* **2019**, *14* (10), 1–16. https://doi.org/10.1371/journal.pone.0224453.

[11] Ali, A.-R.; Li, J.; Yang, G.; O'Shea, S. J. A Machine Learning Approach to Automatic Detection of Irregularity in Skin Lesion Border Using Dermoscopic Images. *PeerJ Comput. Sci.* **2020**, *6*, e268. https://doi.org/10.7717/peerj-cs.268.

[12] Modinos, G.; Mechelli, A.; Pettersson-Yeo, W.; Allen, P.; McGuire, P.; Aleman, A. Pattern Classification of Brain Activation during Emotional Processing in Subclinical Depression: Psychosis Proneness as Potential Confounding Factor. *PeerJ* **2013**, *2013* (1), 1–14. https://doi.org/10.7717/peerj.42.

[13] Lei, L.; Wang, Y.; Xue, Q.; Tong, J.; Zhou, C. M.; Yang, J. J. A Comparative Study of Machine Learning Algorithms for Predicting Acute Kidney Injury after Liver Cancer Resection. *PeerJ* **2020**, *2020* (2). https://doi.org/10.7717/peerj.8583.

[14] Kononenko, I. Machine Learning for Medical Diagnosis: History, State of the Art and Perspective. *Artif. Intell. Med.* **2001**, *23* (1), 89–109. https://doi.org/10.1016/S0933-3657(01)00077-X.

[15] Yves, G. *Bibliometrics and Research Evaluation: Uses and Abuses*; Buckland, M., Furner, J., Krajewski, M., Eds.; MIT Press: Cambridge, MA, 2014.

[16] Aksnes, D. W.; Langfeldt, L.; Wouters, P. Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open* **2019**, *9* (1).

https://doi.org/10.1177/2158244019829575.

[17] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria 2021.

[18] Aria, M.; Cuccurullo, C. Bibliometrix : An R-Tool for Comprehensive Science Mapping Analysis. *J. Informetr.* **2017**, *11* (4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007.

[19] Vickery, B. C. Bradford's Law of Scattering. *J. Doc.* **1948**, *4* (3), 198–203. https://doi.org/10.1108/eb026133.

[20] Cobo, M. J.; López-Herrera, A. G.; Herrera-Viedma, E.; Herrera, F. An Approach for Detecting, Quantifying, and Visualizing the Evolution of a Research Field: A Practical Application to the Fuzzy Sets Theory Field. *J. Informetr.* **2011**, *5* (1), 146–166. https://doi.org/10.1016/j.joi.2010.10.002.

[21] Bejnordi, B. E.; Veta, M.; Van Diest, P. J.; Van Ginneken, B.; Karssemeijer, N.; Litjens, G.; Van Der Laak, J. A. W. M.; Hermsen, M.; Manson, Q. F.; Balkenhol, M.; Geessink, O.; Stathonikos, N.; Van Dijk, M. C. R. F.; Bult, P.; Beca, F.; Beck, A. H.; Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; Zhong, A.; Dou, Q.; Li, Q.; Chen, H.; Lin, H.-J.; Heng, P.-A.; Haß, C.; Bruni, E.; Wong, Q.; Halici, U.; Öner, M. U.; Cetin-Atalay, R.; Berseth, M.; Khvatkov, V.; Vylegzhanin, A.; Kraus, O.; Shaban, M.; Rajpoot, N.; Awan, R.; Sirinukunwattana, K.; Qaiser, T.; Tsang, Y.-W.; Tellez, D.; Annuscheit, J.; Hufnagl, P.; Valkonen, M.; Kartasalo, K.; Latonen, L.; Ruusuvuori, P.; Liimatainen, K.; Albarqouni, S.; Mungal, B.; George, A.; Demirci, S.; Navab, N.; Watanabe, S.; Seno, S.; Takenaka, Y.; Matsuda, H.; Phoulady, H. A.; Kovalev, V.; Kalinovsky, A.; Liauchuk, V.; Bueno, G.; Fernandez-Carrobles, M. M.; Serrano, I.; Deniz, O.; Racoceanu, D.; Venâncio, R.; Consortium, C. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer. *JAMA - J. Am. Med. Assoc.* **2017**, *318* (22), 2199–2210. https://doi.org/10.1001/jama.2017.14585.

[22] Schmidt-Kittler, O.; Ragg, T.; Daskalakis, A.; Granzow, M.; Ahr, A.; Blankenstein, T. J. F.; Kaufmann, M.; Diebold, J.; Arnholdt, H.; Müller, P.; Bischoff, J.; Harich, D.; Schlimok, G.; Riethmüller, G.; Eils, R.; Klein, C. A. From Latent Disseminated Cells to Overt Metastasis: Genetic Analysis of Systemic Breast Cancer Progression. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (13), 7737–7742. https://doi.org/10.1073/pnas.1331931100.

[23] Cruz, J. A.; Wishart, D. S. Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Inform.* **2006**, *2*, 59–77. https://doi.org/10.1177/117693510600200030.

[24] Ganz, P. A.; Hirji, K.; Sim, M.-S.; Schag, C. A. C.; Fred, C.; Polinsky, M. L. Predicting Psychosocial Risk in Patients with Breast Cancer. *Med. Care* **1993**, *31* (5), 419–431. https://doi.org/10.1097/00005650-199305000-00004.

[25] Kilday, J.; Palmieri, F.; Fox, M. D. Classifying Mammographic Lesions Using Computerized Image Analysis. *IEEE Trans. Med. Imaging* **1993**, *12* (4), 664–669. https://doi.org/10.1109/42.251116.

[26] Bakator, M.; Radosav, D. Deep Learning and Medical Diagnosis: A Review of Literature. *Multimodal Technol. Interact.* **2018**, *2* (3). https://doi.org/10.3390/mti2030047.

[27] Yassin, N. I. R.; Omran, S.; El Houby, E. M. F.; Allam, H. Machine Learning Techniques for Breast Cancer Computer Aided Diagnosis Using Different Image Modalities: A Systematic Review. *Comput. Methods Programs Biomed.* **2018**, *156*, 25–45. https://doi.org/10.1016/j.cmpb.2017.12.012.

[28] Litjens, G.; Sánchez, C. I.; Timofeeva, N.; Hermsen, M.; Nagtegaal, I.; Kovacs, I.; Hulsbergen-Van De Kaa, C.; Bult, P.; Van Ginneken, B.; Van Der Laak, J. Deep Learning as a Tool for Increased Accuracy and Efficiency of Histopathological Diagnosis. *Sci. Rep.* **2016**, *6*. https://doi.org/10.1038/srep26286.

[29] Abdel-Zaher, A. M.; Eldeib, A. M. Breast Cancer Classification Using Deep Belief Networks. *Expert Syst. Appl.* **2016**, *46*, 139–144. https://doi.org/10.1016/j.eswa.2015.10.015.

Table **1**. Top ten most cited publications related to breast cancer and machine learning.

Title	**Journal**	**Year**	**Cited**
Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer	JAMA	2017	556
From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression	Proceedings of the National Academy of Sciences of the United States of America	2003	466
Applications of machine learning in cancer prediction and prognosis	Cancer Informatics	2006	431
Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis	Scientific Reports	2016	331
Comprehensive analysis of risk factors associating with hepatitis b virus (HBV) reactivation in cancer patients undergoing cytotoxic chemotherapy	British Journal of Cancer	2004	249
Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms	Expert Systems with Applications	2014	249
A constructive algorithm for training cooperative neural network ensembles	IEEE Transactions on Neural Networks	2003	241
Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer	The American Journal of Surgery	2008	204
Breast cancer classification using deep belief networks	Expert Systems with Applications	2016	204
Volatile markers of breast cancer in the breath	Breast Journal	2003	202

0929-8673/19 $58.00+.00

Table **2**. Top ten countries with the highest total citation.

Country/Region	**Total citation**	**Average article citation**
USA	4629	145
Germany	1546	172
China	1164	129
Netherlands	829	276
Canada	666	222
Spain	420	140
India	339	170
Hungary	304	152
Singapore	287	144
United Kingdom	286	143

Table **3**. Top institution based on a fraction of co-authors per paper.

Institutions	**Country**	**Fraction of co-authors per paper**	**Number of co-authors**	**Number of first authors**
University of Chicago	USA	3.09	6	4
National University of Singapore	Singapore	3.00	3	3
University of Wisconsin	USA	275	8	2
University of Pennsylvania	USA	1.64	4	1
University of Heidelberg	Germany	1.50	2	1
University of Veszprem	Hungary	1.50	2	2
University of Minnesota	USA	1.45	4	2
China Medical College and Hospital	Taiwan	1.33	3	2
Massachusetts Institute of Technology	USA	1.20	4	2
Chinese University of Hong Kong	Hong Kong	1.03	3	1

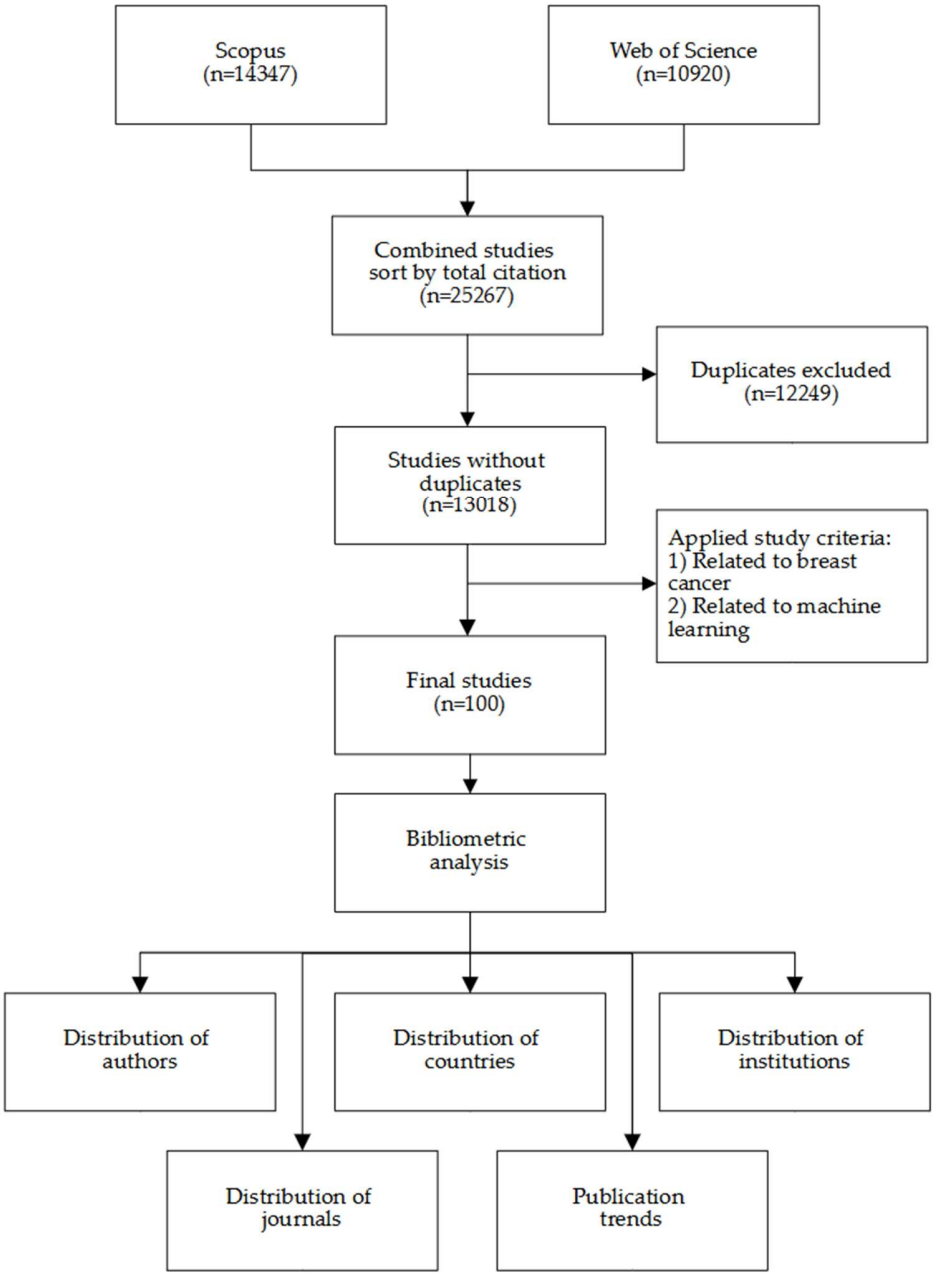

Fig. (1). The flow of study selection and bibliometric analysis.

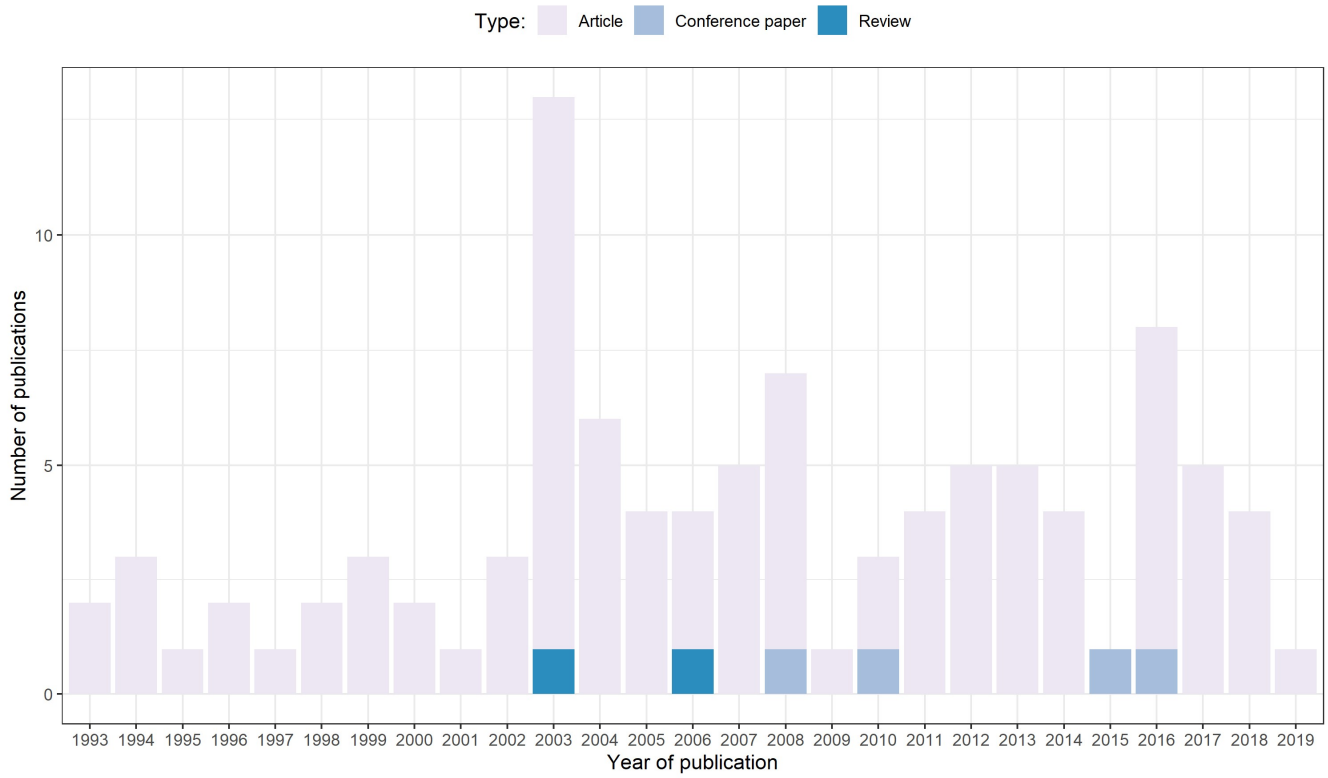

Fig. (2). The distribution of the top 100 most cited publications related to breast cancer and machine learning according to the year of publication.

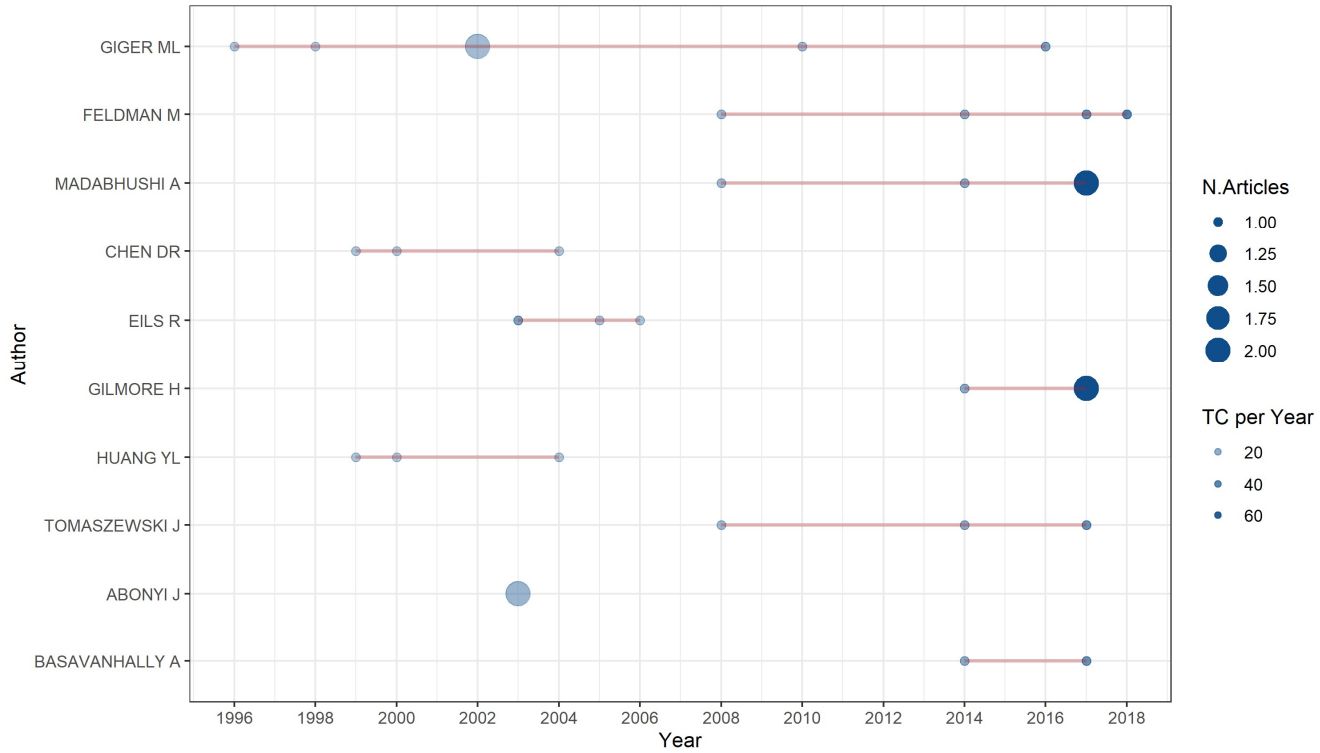

Fig. (3). Top ten productive authors of the top 100 most cited publications related to breast cancer and machine learning.

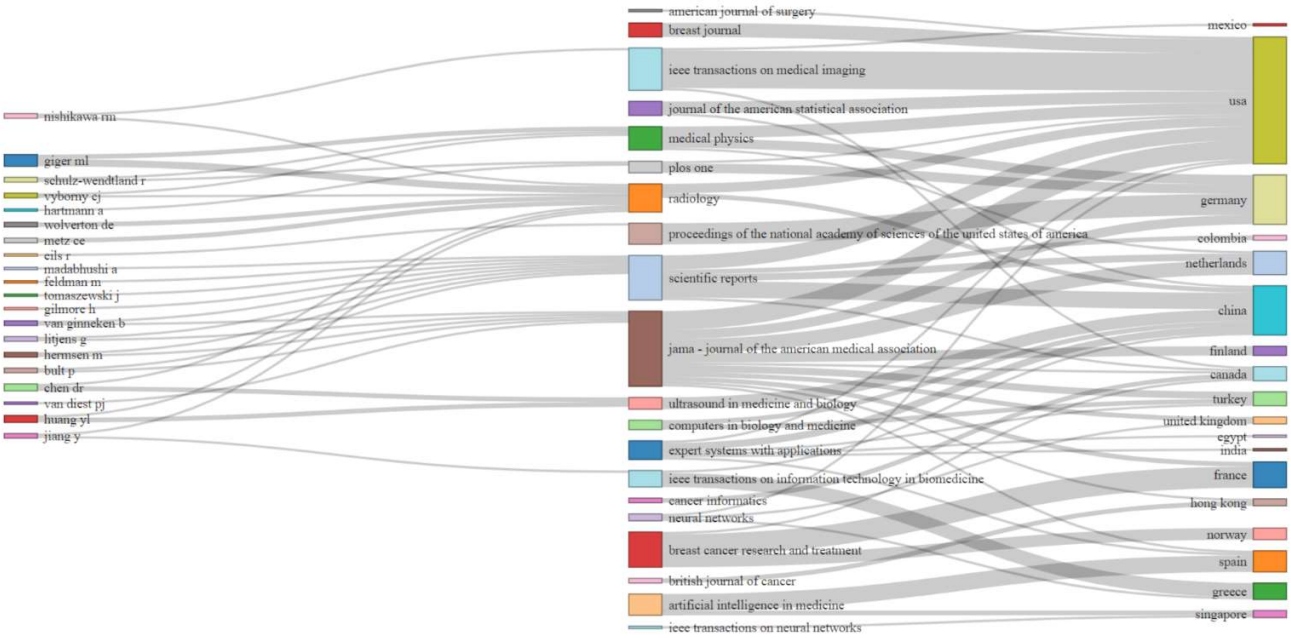

Fig. (4). The three-fields plot for the relationship between top authors (left field), top journals (middle field), and top countries (right field).

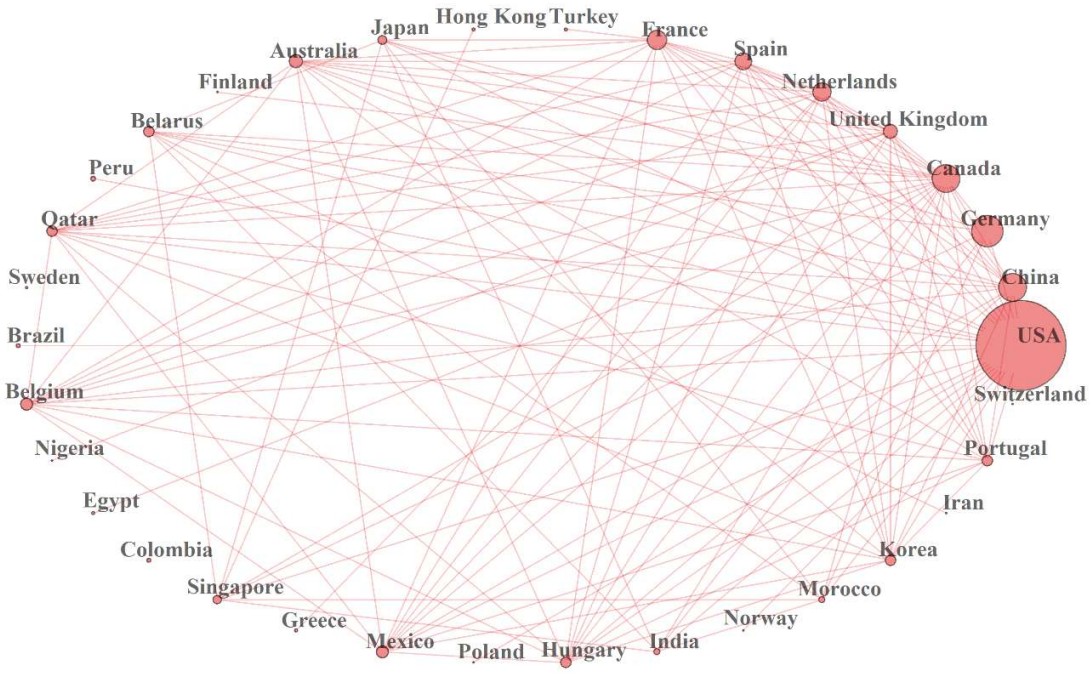

Fig. (5). Countries collaboration in the top 100 most cited publication related to breast cancer and machine learning (Isolated nodes were removed).

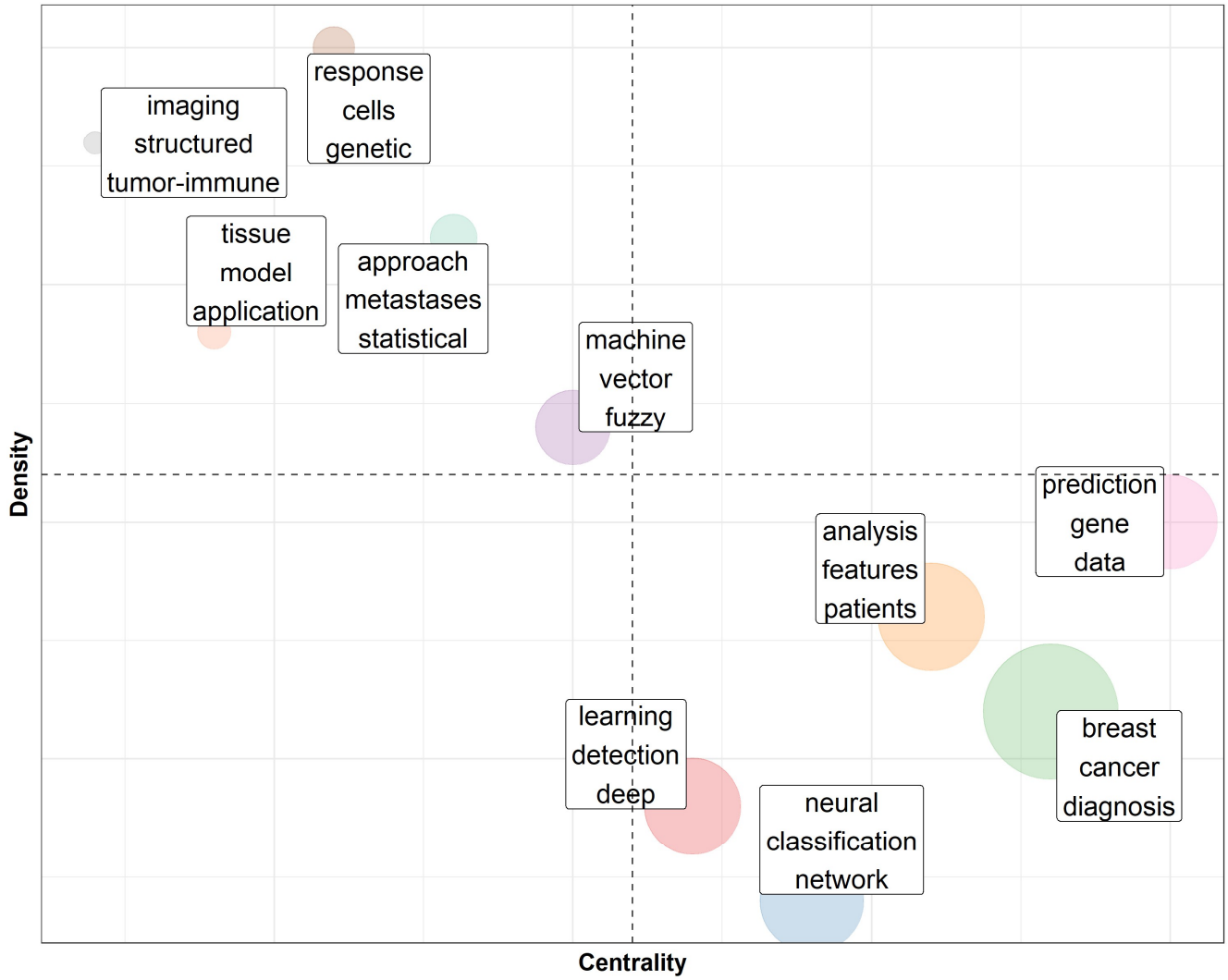

Fig. (6). Thematic map of top 100 most cited publications related to breast cancer and machine learning.

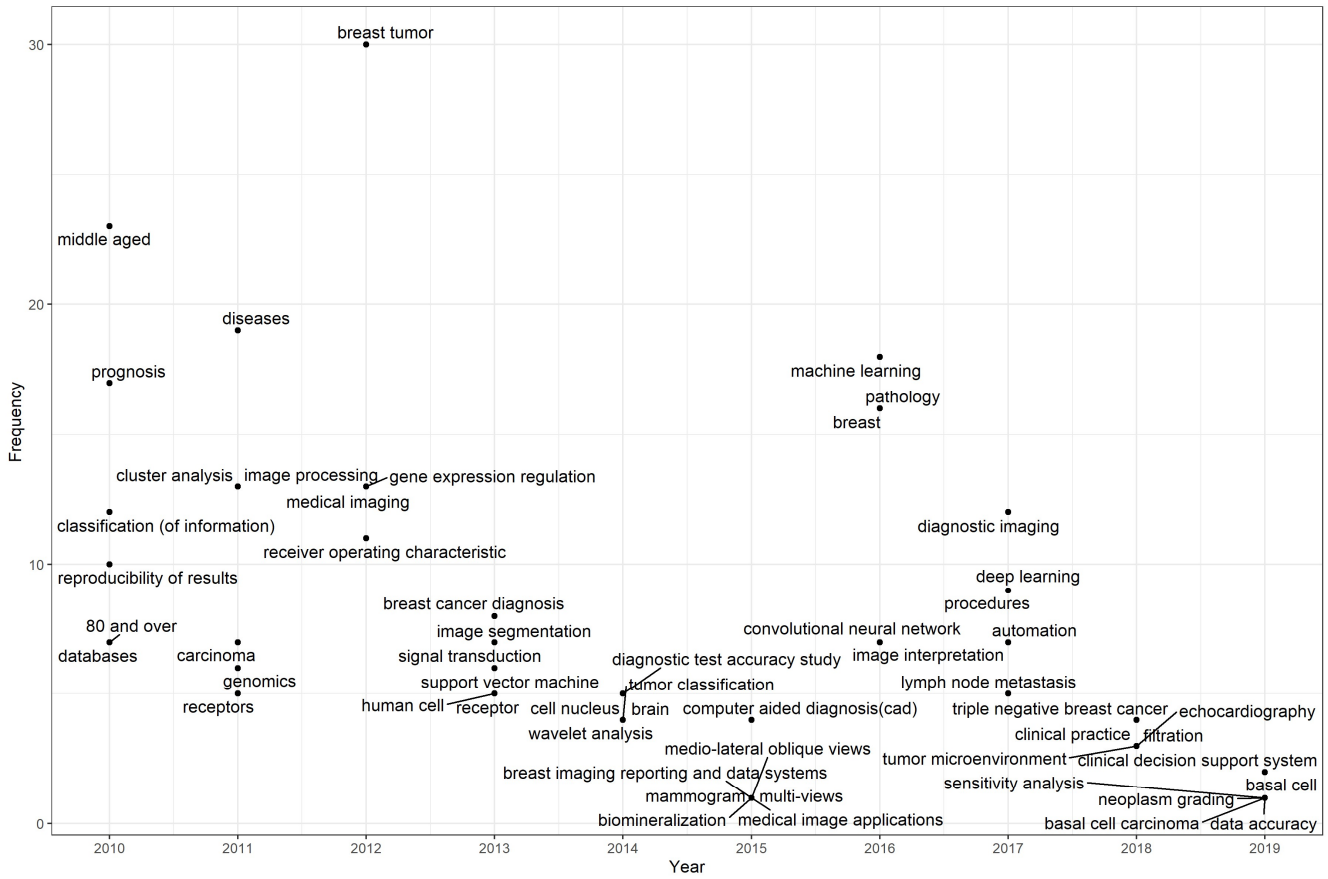

Fig. (7). Trending keywords in the top 100 most cited publication related to breast cancer and machine learning between 2010 and 2019.