



A Comprehensive Bibliometric Analysis of Deep Learning Techniques for Breast Cancer Segmentation: Trends and Topic Exploration (2019-2023)

Agus Perdana Windarto¹, Anjar Wanto², Solikhun³, Ronal Watrianthos⁴

¹Bachelor's Program in Information Systems, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

²Bachelor's Program in Computer Engineering, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

³Bachelor's Program in Information Management, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

⁴Bachelor's Program in Computer Engineering, Universitas Al Washliyah, Rantauprapat, Indonesia

¹agus.perdana@amiktunasbangsa.ac.id, ²anjarwanto@amiktunasbangsa.ac.id, ³solikhun@amiktunasbangsa.ac.id,

⁴ronal.watrianthos@gmail.com

Abstract

The objective of this study is to perform a comprehensive bibliometric analysis of the existing literature on breast cancer segmentation using deep learning techniques. Data for this analysis were obtained from the Web of Science Core Collection (WOS-CC) that spans from 2019 to 2023. The study is based on a comprehensive collection of 985 documents that cover a substantial body of research findings related to the application of deep learning techniques in segmenting breast cancer images. The analysis reveals an annual increase in the number of published works at a rate of 16.69%, indicating a consistent and robust increase in research efforts during the specified time frame. Examining the occurrence of keywords from 2019 to 2023, it is evident that the term "convolutional neural network" exhibited a notable frequency, reaching its peak in 2021. However, the term "machine learning" demonstrated the highest overall frequency, peaking around 2021 as well. This emphasizes the importance of machine learning in the advancement of image segmentation algorithms and convolutional neural networks, which have shown exceptional effectiveness in image analysis tasks. Furthermore, the utilization of latent Dirichlet Allocation (LDA) to identify topics resulted in a relatively uniform distribution, with each topic having an equivalent number of abstracts. This indicates that the data set encompasses a diverse range of topics within the field of deep learning as it relates to breast cancer image segmentation. However, it should be noted that topic 4 has the highest level of significance, suggesting that the application of deep learning for diagnosis was extensively explored in this study.

Keywords: breast cancer; deep learning; image segmentation; bibliometric; WOS

1. Introduction

Breast cancer is a significant and potentially lethal medical condition [1]. Detecting breast cancer in a timely and precise manner is of utmost importance to ensure optimal treatment outcomes, given its major role as a prominent contributor to mortality rates among women worldwide [2]. To facilitate the diagnosis and formulation of treatment strategies for breast cancer, it is imperative to accurately distinguish and distinguish malignant regions from ultrasound images [3].

Deep learning algorithms have shown remarkable abilities in the analysis of medical images, specifically in the segmentation of breast cancer from ultrasound [4] - [12] and mammography images [13]-[16]. Precise segmentation plays a crucial role in the

identification of tumor boundaries, the evaluation of tumor dimensions, and the provision of essential data for the stratification of treatment [17].

Deep learning is an artificial intelligence technique that enables computers to learn and recognize complex patterns in data [18], [19]. In the context of breast cancer detection, deep learning is used to analyze ultrasound images to identify and separate cancer areas with high precision [20]. The advantage of this approach lies in its ability to handle the complexity and variability of medical images, allowing for more accurate and efficient breast cancer detection [21].

Despite increasing interest and advances in deep learning-based breast cancer segmentation, a systematic review of existing research is needed. This will help to understand the current state of affairs in the field, identify emerging trends, and highlight areas

that require further investigation. The bibliometric review of this study will examine the literature on the role of deep learning in accurate breast cancer.

To achieve research objectives, a comprehensive analysis of research articles indexed in the Web of Science database (2018–2023) will be carried out to examine the literature on deep learning-based breast cancer segmentation. A bibliometric study was conducted from 2019 to 2023 using the Web of Science Core Collection (WOS-CC) database. In this study, the researchers used specific keywords, such as 'breast cancer', for the title and 'deep learning' on the topic (title, abstract, keyword). By systematically examining the existing literature, this study will identify relevant articles, influential authors, and trends in the use of deep learning in the segmentation of the breast cancer image. The results of this bibliographic analysis will provide valuable information to researchers, educators, and policy makers interested in applying deep learning to image segmentation. However, it is important to note that this bibliometric study has limitations, including the size of the database and the keywords used. Although this research provides insight into advances in breast cancer research using deep learning and segmentation techniques, it is important to consider its limitations.

2. Research Methods

The method used in this research involves performing a bibliometric analysis using the Web of Science Core Collection (WOS-CC) database [22]. The impetus to use the Web of Science Core Collection (WOS-CC) as the singular data source for this bibliometric analysis covering 2019-2023 was its substantive scholarly coverage and trusted citation data, allowing for a rigorous systematic study. Specifically, the WOS-CC includes more than 21,000 journals that span scientific disciplines and incorporates citation links between publications, allowing for enriched bibliometric mapping [23].

Furthermore, it constitutes a curated aggregation of sources deemed to impart significant scholarly impact and influence. Therefore, focusing our search within the WOS-CC provided an in-depth analysis of the high-quality literature integral to our selected domain of deep learning for breast cancer segmentation. Although integrating additional databases could surface worthwhile vantage points, we posit the WOS-CC furnished an appropriately rigorous lens to fulfill our aims for this focused bibliometric study given current resource constraints. We invite discussion of both the advantages and limitations of our chosen approach.

Bibliometrics is the systematic and quantitative analysis of scientific publications and the citations they contain [24], [25]. The use of this tool has been

proven to be extremely beneficial in monitoring the progress of research areas, identifying emerging patterns, and evaluating the influence exerted by individual researchers and institutions [26], [27].

The WOS-CC database is considered to be the most extensive and most frequently used bibliometric database worldwide [28]. This comprehensive database includes a collection of more than 21,000 scientific journals from various academic fields, as well as conference proceedings and books. In addition to its main content, the database contains citation data that allows researchers to monitor the impact and reception of their scientific contributions over a period of time [29].

Although there are other databases such as Scopus that offer valuable features and extensive coverage, the choice of WOS for this study was motivated by its clear advantages [29]. WOS is recognized as the most comprehensive, trusted, and user-friendly bibliometric database available today. This tool is considered very suitable for performing bibliometric analyzes in the context of research fields [28].

Several factors play a role in influencing the reliability and quality of research findings obtained from the WOS database. The analysis was carried out in July 2023, using precise search queries to retrieve pertinent documents. The search query used was "Breast Cancer" (title) and "Deep Learning" (topic) and 2023 or 2022 or 2021 or 2020 or 2019 (publication years) and Article or Proceeding Paper (Document Types) and Article or Proceeding Paper (Document Types), with the aim of identifying documents that contained the term "Breast Cancer" in their title and were classified under the topic of "Deep Learning".

The purpose of this bibliometric analysis is to examine and evaluate the scientific literature on the application of deep learning algorithms in the segmentation of breast cancer images. The main objective of the researcher is to gain a comprehensive understanding of the existing research on the subject. After completion of the data retrieval process, a total of 985 documents were captured, all of which met the specified search criteria. The above documents are used as a basis for subsequent review and evaluation, providing a valuable point of view on patterns, motivations, and advances in this particular area. This study uses Biblioshiny as a data visualization tool that facilitates descriptive analysis and conceptual testing [30].

3. Results and Discussions

The bibliometric analysis presented in this section aimed to fulfill three main objectives. The first was to quantify publication output and growth in order to elucidate research trends over the specified time period. The second objective was to identify the

impactful contributions and contributors to chart the principal investigators advancing knowledge in the field. Finally, the third goal was to elucidate the predominant topics by inspecting keywords and abstract text to understand the primary investigative foci.

The results delineated in the following sections are structured to address each of these objectives and provide a comprehensive perspective on the landscape of the literature. Elucidating these motivations will facilitate the interpretation of the forthcoming analysis and demonstrate how the different bibliometric components connect to the overarching study goals. Guided by this direction, we present the findings obtained by our systematic methodology.

The data set provided includes an extensive bibliometric analysis performed on the topic of deep learning in breast cancer image segmentation. The analysis spans five years, from 2019 to 2023, and provides a comprehensive overview of the scholarly discourse on the topic during this period.

The richness of the data is due to the different sources. The information was compiled from 461 unique sources, including scientific journals, books, and other relevant publications. This diverse range of sources allows for a comprehensive understanding of the subject and encompasses various perspectives and methods used in this research area.

A total of 985 documents form the basis of this study and represent a significant body of research in the field of deep learning applications in breast cancer image segmentation. This large number of documents illustrates the dynamism of the field and the great attention it received from the scientific community. In particular, the annual growth rate of the number of publications is 16.69%, indicating a steady and robust increase in research activity over the given period. This growth rate reflects the growing interest and investment in this area and the recognition of its potential to advance the diagnosis and treatment of breast cancer.

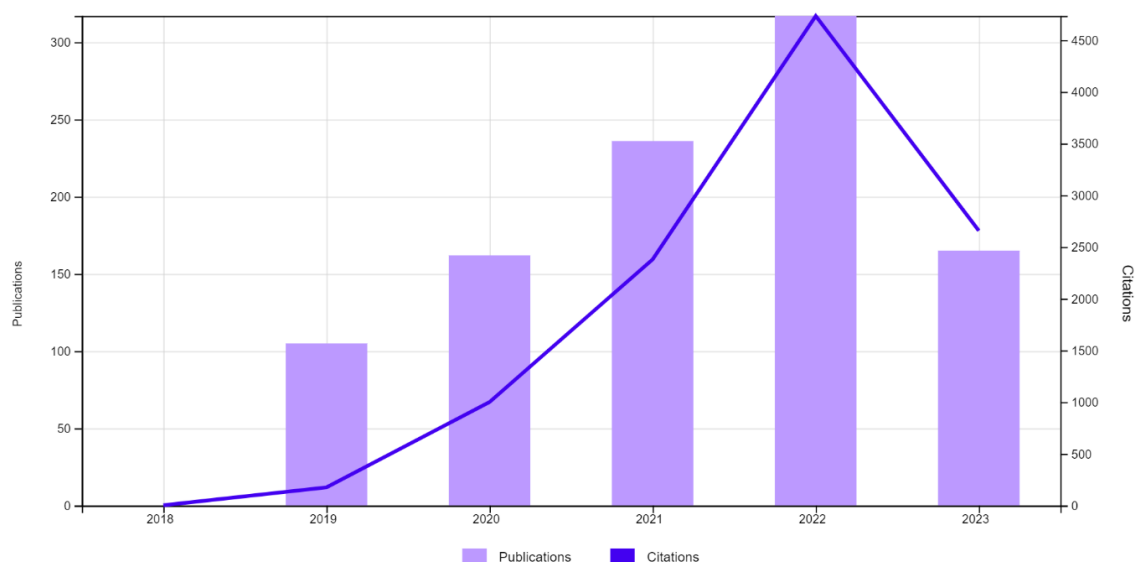


Figure 1. Time Cited and Publication (2019-2023)

Figure 1 shows the annual scientific output in deep learning in breast cancer image segmentation from 2019 to 2023. In 2019, the research community produced 103 articles on this topic. This production increased significantly in 2020, and the number of items increased to 157. The upward trend continued in 2021 with a further increase to 210 articles.

The most notable increase in scientific production occurred in 2022, when the number of articles reached 324. By mid-2023, the number of articles had already reached 191 when this dataset was withdrawn from WOS-CC. This indicates a strong and growing interest in applying deep learning techniques to segment breast cancer images.

The data set provides information on the total number of annual citations per year from 2019 to 2023 for deep learning in breast cancer image segmentation. In 2019, each published article received an average of 36.37 total citations. Since the articles have been in circulation for five years, they have been cited an average of 7.27 times per year.

In 2020, the average total number of citations per article fell slightly to 20.47, which is an average of 5.12 citations per year given the four years these articles were available. This downward trend in the average number of citations per article and the average number of citations per year continued in subsequent years. In 2021, the average total number of citations

per article was 11.13 and the average number of citations per year was 3.71. In 2022, these numbers dropped to 4.85 and 2.42.

Finally, in 2023, the average total number of citations per article was 0.43, which corresponds to the average number of citations per year assuming that the articles had only been available for citation for a year. An important note is that a decrease in citations per article and per year does not necessarily indicate a decrease in impact or interest in the field. Recent published articles are cited less often because they have less time to be cited in other works. However, as can be seen in Figure 1, the trend of citations continues to increase each year according to the trend of publication.

3.1 Main Information

The application of deep learning in the segmentation of breast cancer images has been the subject of intensive scientific research over a 5-year period, from 2019 to 2023. This period provides a comprehensive and meaningful overview of the scientific dialogue, emerging trends, and key breakthroughs that shaped the field during this period. The data set used for this bibliometric analysis is characterized by its diversity and breadth. The data used consisted of 461 unique sources summarizing a wide range of relevant scientific journals, books, and other scientific publications.

When examining the most influential sources, Frontiers in Oncology emerged as the most significant author with a total of 33 articles. Among them was Diagnostics with an extensive contribution of 30 articles. IEEE Access and Cancer also made significant contributions with 29 and 26 articles, respectively. During the five years from 2019 to 2023, the steady flow of articles from these and other sources underscores the importance of deep learning in the segmentation of breast cancer images.

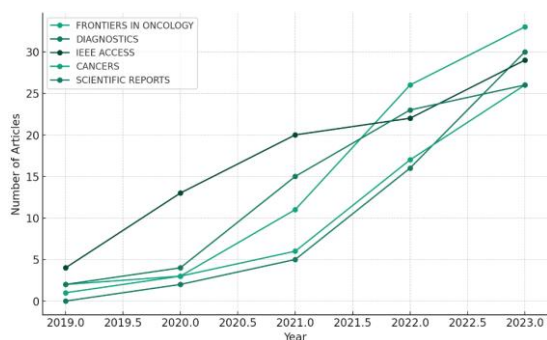


Figure 2. Sources' Production Over Time (2019-2023)

Figure 2 illustrates the dynamics of article production from five main sources of deep learning in breast cancer image segmentation from 2019 to 2023. Frontiers in Oncology shows a significant increase in the number of articles over the years, starting with just

one article in 2019 and a sharp increase to 33 articles in 2023. This strong growth underscores the role of the journal as the leading platform for the dissemination of research in the field.

Diagnostics also saw steady growth, from no articles in 2019 to 30 articles in 2023. IEEE Access also saw a notable increase in the number of articles published, starting with 4 articles in 2019 and peaking at 29 articles in 2023. The Cancer and Science Reports show significant growth in the number of articles over the period. Cancer started with 2 articles in 2019 and reached 26 articles in 2023, while Scientific Reports went from 2 articles in 2019 to 26 articles in 2023.

Table 1. The Most Contributed Authors

Authors	Record Count	% Of 985
Wang Y	17	1.72
Liu Y	11	1.11
Zhang L	11	1.11
Idri A	9	0.91
Khan MA	9	0.91
Li J	9	0.91
Sarkar R	9	0.91
Zerouaoui H	9	0.91
Liu ZY	8	0.81
Wang J	8	0.81

Table 1 shows that 10 authors are among the 4,561 authors who contributed to this area during this period. Wang Y is the most prolific author in this field, with a record 17, contributing about 1.726% of the 985 total documents. However, this contribution is relatively small compared to the total number of documents found. This means that the author contributions are fairly evenly distributed and that no one dominates in this study.

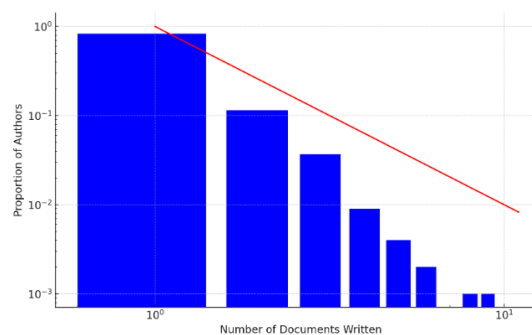


Figure 3. Author Productivity Through Lotka's Law (2019-2023)

Figure 3 shows the distribution of deep learning author productivity in breast cancer image segmentation from 2019 to 2023 as described by Lotka's law [31]. The blue bars represent the actual data, while the red line represents the theoretical distribution according to Lotka's law. Lotka's law is a principle that describes the frequency of publications by authors in a given field. The authors of the number of authors making n contributions, which is approximately $1/n^2$ of those

making a contribution. This law is often used to study the distribution of scientific productivity.

The red line in the diagram shows the theoretical curve predicted by Lotka's law. As we can see, the actual distribution of author productivity (blue bars) generally follows the trend predicted by Lotka's law. Most authors (83.2%) have only written one article in this area, while the proportion of authors who have written two documents is 11.4%. There are also 3.7% of the authors who wrote three documents. The proportion of authors continues to decrease as the number of documents written increases. These differences could be due to a variety of factors, including collaboration patterns, the interdisciplinarity of the subject studied, or other characteristics of the research community.

However, analyzing author productivity using Lotka's law for deep learning in breast cancer image segmentation shows a pattern consistent with many other research areas: a large number of authors contribute a small number of works, while a small number contribute high-quality works. This pattern underscores the diversity and dynamism of the research community in this area.

As the most prolific author in the field, Wang's research on image segmentation of breast cancer using deep learning has the potential to improve early detection and diagnosis of breast cancer. By automatically segmenting breast tumors into MRI images, the method can help radiologists identify and assess tumors with greater precision. This could lead to earlier diagnosis and treatment of breast cancer, which could improve patient outcomes. With a significant number of widely cited publications, Wang's work had a significant impact on this area of research. In the field of deep learning in breast cancer image segmentation, Y. Wang has made significant scientific contributions, as evidenced by several important bibliometric indicators.

Y. Wang has an h-index of 6, a value that speaks for both productivity and citation effectiveness. In context, this means that six of Y. Wang's publications have been cited at least six times in other works, underscoring the impact and reach of her research. In addition, Y. Wang's g-index is 17. As an alternative to the h index, the g index places additional emphasis on frequently cited articles. In the case of Y. Wang, this means that her 17 most cited works have a total of at least 289 citations.

Although Y. Wang is the most prolific author of this study, Shen et al. (2019) is the most influential publication on deep learning in breast cancer image segmentation based on the WOS-CC database. The article entitled 'A deep learning mammography-based model for improved breast cancer risk prediction' was

edited by Yala, Adam; Lehman, Constance; Cobbler, Valley; Portnoi, Tally; and Barzilay, Regina. This research, published in *Radiology* in July 2019, represents a significant contribution in this field with a total of 310 citations [32].

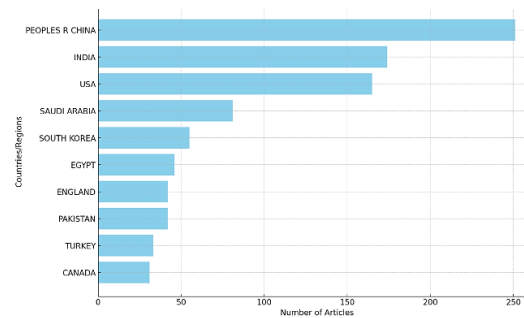


Figure 4. Top Ten Countries/Region by Number Articles

The high total number of citations indicates that the study had a strong impact on the field and was widely cited by other researchers. While the average number of citations is 62, this article significantly exceeded the average, underscoring its impact and relevance in the research community.

The dominance of Chinese authors in this study can also be seen in Figure 4, the country distribution of the authors of the 985 articles in this data set related to research on the use of deep learning in breast cancer image segmentation. The authors of the People's Republic of China contributed the most articles with 251 articles, accounting for 25.482% of the total. This indicates a significant level of research activity in these countries.

3.2 Discussions

Deep learning has been widely applied to breast cancer image segmentation in the last five years, from 2019 to 2023. This approach has been used to detect, segment, extract features, and classify tumors, leading to advanced results. It is also applied to almost all breast cancer imaging tests, including MRI [33].

A study introduced a flexible, unsupervised deep learning model called Divide-and-Conquer (dc)-DeepMSI that leverages the divide-and-conquer strategy to facilitate design and application for analyzing metabolic heterogeneity from mass spectrometry-imaging data (MSI) to develop without prior knowledge of histology. This model can identify spatially contiguous regions of interest (ROI) or spatially sporadic ROIs by designing two specific modes: late-contiguous and late-spor [34].

Deep learning algorithms have been used for segmenting breast cancer images with promising results. Convolutional neural networks (CNN) are a type of deep learning algorithm that has been shown to be effective in segmenting tumors in breast cancer images. CNNs work by learning to recognize patterns

in images and can be trained to segment tumors with high precision [18].

Another deep learning algorithm that has been used to segment breast cancer images is the grasshopper optimization algorithm (GOA). GOA is a metaheuristic algorithm that can be used to optimize the parameters of CNNs. By optimizing the parameters of the CNNs, GOA can improve the performance of CNNs for breast cancer image segmentation [35].

Algorithms such as dc-DeepMSI are specifically designed for breast cancer. dc-DeepMSI uses a combination of CNNs and Markov random fields to segment tumors in breast cancer images. dc-DeepMSI has been shown to be effective in segmenting tumors in breast cancer images and was found to be more accurate than other deep learning algorithms for segmenting breast cancer images [36].

The deep learning algorithm has shown promising results in segmenting breast cancer images. The annual growth rate of related publications reached 16.69% in the period 2019-2023, indicating a steady and strong increase in research activity during this period.

Frontiers in Oncology is proving to be a leading academic platform when examining key sources contributing to the field of deep learning in breast cancer image segmentation. With a total of 33 published articles in this specialty, the journal has advanced the discourse and created significant value for the research community.

One of the articles that stood out for its profound impact was the study published in 2020 by Sun et al. [37]. This research article examined the comparison between deep convolutional neural network (CNN) models and radiomic analysis to predict axillary lymph node (ALN) metastasis. This study found that CNNs showed numerically superior overall performance in predicting ALN metastases in breast cancer compared to radiomic models. The use of deep learning techniques, particularly convolutional neural networks (CNN), can improve the accuracy of predicting axillary lymph node metastasis (ALN) in patients with

breast cancer using ultrasound images. When intra- and peritumoral regions are seen on ultrasound images, the precision of the prediction model can be significantly improved [37].

In another study, a deep learning algorithm was able to accurately detect breast cancer using a training data set with full clinical annotations or just the cancer status from the entire image, efficiently using an end-to-end training approach for mammogram screening. Shen et al. (2019) examine this in 'Deep Learning to Improve Breast Cancer Detection in Mammography Screening', published in *Scientific Reports*, Volume 12496 (2019) [38].

This article demonstrates the potential of deep learning methods to improve the accuracy and efficiency of breast cancer screening and diagnosis. The developed deep learning algorithms achieve excellent performance compared to previous methods, and the transferability of the entire image classifier trained to mammograms with an end-to-end approach is promising.

The authors hypothesize that with increasing training datasets and available computational resources, deep learning methods have the potential to further improve the precision of breast cancer detection in mammography screening. This finding makes this article the most influential of all articles, with a total of 310 citations in WOS-CC during the study period [38].

Research by Sun et al. (2020) and Shen et al. (2019) shows that the convolutional neural network (CNN) is the most widely used algorithm in image segmentation. CNN is very well suited for image segmentation because it can learn to recognize patterns in the image and can be trained to segment objects with high accuracy [39]. CNNs can be trained to segment objects by feeding them a data set of manually segmented images [40]. CNN learns to identify patterns associated with each object in the data set and can then be used to segment new images containing the same objects [41].

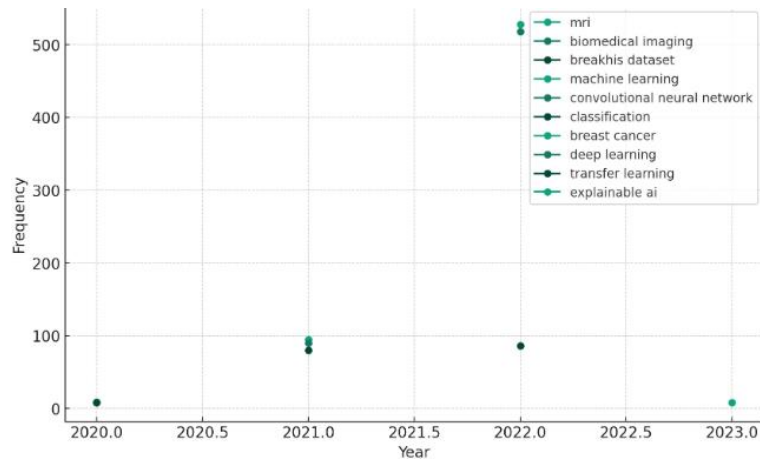


Figure 5. Trend of Keyword Occurrences (2019-2023)

In the trend of annual keyword occurrence from 2019 to 2023 in deep learning in breast cancer image segmentation (Figure 5), the convolutional neural network shows high frequency, with a peak in 2021. Machine learning shows the highest frequency with a peak around 2021. The importance of machine learning in the development of image segmentation algorithms and CNN that are particularly effective for image analysis tasks underscores the importance of these techniques in this area [42] [43].

The event analysis presented in Figure 5 also shows the keyword MRI (Magnetic Resonance Imaging) and shows a stable frequency of occurrence of this keyword over the years. This suggests that magnetic resonance images are a common type of data used in this field to develop and test image segmentation algorithms [44]. Likewise, the keyword 'biomedical imaging' indicates that the broader field of biomedical imaging provides the main context for research in this area. Of particular interest is the record keyword BreakHis, which peaks in frequency around 2020.

The BreakHis dataset is a special dataset of histopathological images of breast tumors that is widely used in research to develop and test image segmentation algorithms [45]. The use of specific data types, such as magnetic resonance imaging and the BreakHis data set, underscores the practical and application-oriented nature of research in this area [46].

The trend in keyword occurrence shown in Figure 5 is based only on the author's keywords in each article. Therefore, we have added topic identification to identify the main topics or topics covered in the articles. This can provide a general overview of this research area and identify important areas of focus.

We applied the Latent Dirichlet Allocation (LDA) model [47], a type of probabilistic topic model that assigns topics to documents and words to topics in a way that balances the breadth and specificity of topics. This approach may be more meaningful for text data, as it can reveal the themes underlying the abstracts [48].

Table 2. Topic Identification from Abstract Using LDA

Topic	Top 10 Keywords	Inferred Topic
Topic 1	for, is, learning, images, on, model, dataset, classification, models, with	Learning models using image datasets for classification
Topic 2	is, accuracy, are, for, this, images, proposed, detection, on, segmentation	Proposed Image Detection and segmentation methods and their accuracy
Topic 3	learning, is, features, machine, deep, are, as, for, with, this	Deep machine learning techniques and features
Topic 4	is, for, this, learning, deep, that, diagnosis, are, with, on	Deep learning applications for diagnosis
Topic 5	for, we, data, with, on, is, that, patients, from, deep	Using deep learning on patient data
Topic 6	for, with, were, was, patients, on, segmentation, as, ct, used	Segmentation techniques used for CT scans on patients
Topic 7	we, images, for, on, classification, model, that, is, image, with	Using Models for image classification
Topic 8	were, with, was, model, for, patients, from, 95, auc, risk	Risk and AUC metrics for models in patient data
Topic 9	model, were, for, patients, was, from, with, prediction, nac, clinical	Predictive modeling using clinical patient data
Topic 10	is, model, for, classification, proposed, feature, with, image, optimization, algorithm	Proposed Optimization Algorithms for image classification models

After a basic text preprocessing approach that includes lowercase, punctuation removal, and tokenization, we used Latent Dirichlet Allocation (LDA) to identify 10 topics from the article abstract. Table 2 shows the top ten keywords for each topic that can contribute to understanding the main theme of each topic. The LDA model identified these themes based on patterns and associations between words in the document corpus. The top 10 keywords are the ten most relevant words or terms associated with each topic. These words have been found to be the most common or most important in the context of the topic in question. These keywords are interpreted as topics or research topics. For example, topic 1's top keywords include for, is, learning, image, on, model, dataset, classification, model, with, indicating that this topic may be about a learning model that uses image datasets for classification.

From the histogram in Figure 6 we can see that the topic distribution is relatively uniform, with each topic having the same number of possible abstracts. This shows that this data set covers a wide range of topics in deep learning in breast cancer image segmentation. However, topic 4 seems to be the most important topic, so deep learning applications for diagnosis were the most examined in this study.

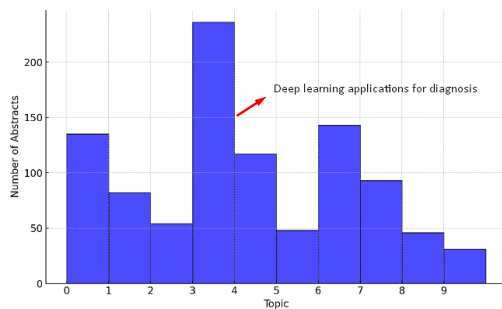


Figure 6. Distribution of Topics in Abstracts

This analysis provides a general overview of the key issues covered in the abstracts. However, it is important to note that the assignment of abstracts to topics is based on a probabilistic model and each abstract is likely to touch on multiple topics to some degree. Thorough reading and analysis are required for a deeper understanding of the abstract content.

4. Conclusions

Accurate segmentation of breast cancer from ultrasound images plays an important role in diagnosis and treatment planning. In recent years, deep learning techniques have received a lot of attention due to their potential to achieve accurate segmentation results. The aim of this study is to provide a comprehensive bibliometric review of the literature related to deep learning-based breast cancer segmentation using Web of Science from 2019 to 2023. A total of 985

documents form the basis of this study and represent a large number of document research results in the field of deep learning applications in breast cancer image segmentation. The annual growth rate of the number of publications reached 16.69%, indicating a steady and strong increase in research activity during the specified period. Frontiers in Oncology proved to be the most important publisher with a total of 33 articles. Authors from the People's Republic of China contributed the most, with 251 articles, accounting for 25.482% of the total number of articles. This indicates a significant level of research activity in the country. Shen et al. (2019) in "Deep Learning to Improve Breast Cancer Detection in Mammography Screening" became the most influential of all articles with a total of 310 citations in WOS-CC during the study period. In the annual trend of keyword occurrences from 2019 to 2023, the keyword convolutional neural network showed a high frequency with a peak in 2021. Machine learning shows the highest frequency with a peak around 2021. The importance of machine learning is high in the development of effective image segmentation algorithms and artificial neural networks for image analysis tasks, highlighting the importance of these techniques in this field. We also analyzed topic identification using latent Dirichlet Allocation (LDA). The result was a relatively even distribution of topics, with each topic containing the same number of abstracts. This shows that this data set covers a wide range of topics in deep learning in breast cancer image segmentation. However, topic 4 seems to be the most important topic, making the application of deep learning to diagnosis the most studied topic of this research. In general, this research provides valuable information on the current state of research related to deep learning-based breast cancer segmentation. The results of this study can guide future research in this area and help improve the precision of breast cancer diagnosis and treatment planning.

Acknowledgment

We express our heartfelt gratitude to the Directorate of Research, Technology, and Community Engagement for their generous funding of our fundamental basic research in 2023. Their financial support has been instrumental in the success of this study and has enabled us to make significant contributions to the field of science. We also extend our sincere appreciation to the evaluation committee for recognizing the value and potential of our research project and providing us with this funding opportunity. Their confidence in our work has been truly motivating. Furthermore, we are extremely grateful to the dedicated members of our research team who have invested their time, expertise, and efforts into this project. Their commitment and hard work have been essential in achieving our research goals. Additionally,

we would like to thank all participants and organizations who contributed to data collection and provided valuable insights and resources for our study. Lastly, we want to acknowledge the support and guidance we have received from our colleagues, mentors, and friends throughout this research endeavor. Their encouragement and constructive feedback have greatly shaped the outcome of this study. We recognize that this research would not have been possible without the collective support and collaboration of all involved. The authors are sincerely grateful for the opportunities and assistance provided, and remain committed to advancing scientific knowledge in our field.

References

- [1] N. Harbeck *et al.*, *Breast cancer*, vol. 5, no. 1, 2019. doi: 10.1038/s41572-019-0111-2.
- [2] L. Ellis, L. M. Woods, J. Estève, S. Elooranta, M. P. Coleman, and B. Rachet, "Cancer incidence, survival and mortality: Explaining the concepts," *Int J Cancer*, vol. 135, no. 8, pp. 1774–1782, 2014, doi: 10.1002/ijc.28990.
- [3] R. M. Reilly, "Ultrasound imaging technology," *Medical Imaging for Health Professionals: Technologies and Clinical Applications*, vol. 44, no. 1, pp. 107–113, 2019, doi: 10.1002/9781119537397.ch6.
- [4] A. A. Ardakani, A. Mohammadi, M. Mirza-Aghazadeh-Attari, and U. R. Acharya, "An open-access breast lesion ultrasound image database: Applicable in artificial intelligence studies," *Comput Biol Med*, vol. 152, 2023, doi: 10.1016/j.compbiomed.2022.106438.
- [5] Q. Q. He, Q. J. Yang, and M. H. Xie, "HCTNet: A hybrid CNN-transformer network for breast ultrasound image segmentation," *Comput Biol Med*, vol. 155, 2023, doi: 10.1016/j.compbiomed.2023.106629.
- [6] H. Meng *et al.*, "DGANET: A Dual Global Attention Neural Network For Breast Lesion Detection In Ultrasound Images," *Ultrasound Med Biol*, vol. 49, no. 1, pp. 31–44, 2023, doi: 10.1016/j.ultrasmedbio.2022.07.006 WE - Science Citation Index Expanded (SCI-EXPANDED).
- [7] S. Y. Lu, S. H. Wang, and Y. D. Zhang, "BCDNet: An Optimized Deep Network for Ultrasound Breast Cancer Detection," *IRBM*, vol. 44, no. 4, 2023, doi: 10.1016/j.irbm.2023.100774.
- [8] Y. Alzahrani and B. Boufama, "Medical-Network (Med-Net): A Neural Network for Breast Cancer Segmentation in Ultrasound Image," *Advanced Intelligent Virtual Reality Technologies, AIVR 2022*, vol. 330, no. 6th International Conference on Artificial Intelligence and Virtual Reality (AIVR), Univ Windsor, 401 Sunset Ave, Windsor N9B 3P4, ON, Canada, pp. 145–159, 2023. doi: 10.1007/978-981-19-7742-8_12 WE - Conference Proceedings Citation Index - Science (CPCI-S).
- [9] M. A. Hassanien, V. K. Singh, D. Puig, and M. Abdel-Nasser, "Predicting Breast Tumor Malignancy Using Deep ConvNeXt Radiomics and Quality-Based Score Pooling in Ultrasound Sequences," *DIAGNOSTICS*, vol. 12, no. 5, 2022, doi: 10.3390/diagnostics12051053 WE - Science Citation Index Expanded (SCI-EXPANDED).
- [10] H. M. Balaha, M. Saif, A. Tamer, and E. H. Abdelhay, "Hybrid deep learning and genetic algorithms approach (HMB-DLGAHA) for the early ultrasound diagnoses of breast cancer," *Neural Comput Appl*, vol. 34, no. 11, pp. 8671–8695, 2022, doi: 10.1007/s00521-021-06851-5.
- [11] W. Q. Shen *et al.*, "Using an Improved Residual Network to Identify PIK3CA Mutation Status in Breast Cancer on Ultrasound Image," *Front Oncol*, vol. 12, 2022, doi: 10.3389/fonc.2022.850515 WE - Science Citation Index Expanded (SCI-EXPANDED).
- [12] X. L. Qu *et al.*, "A VGG attention vision transformer network for benign and malignant classification of breast ultrasound images," *Med Phys*, vol. 49, no. 9, pp. 5787–5798, 2022, doi: 10.1002/mp.15852.
- [13] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Sci Rep*, vol. 9, no. 1, pp. 1–12, 2019, doi: 10.1038/s41598-019-48995-4.
- [14] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A deep learning mammography-based model for improved breast cancer risk prediction," *Radiology*, vol. 292, no. 1, pp. 60–66, 2019, doi: 10.1148/radiol.2019182716.
- [15] M. Siddique, M. Liu, P. Duong, and S. Jambawalikar, "Deep Learning Approaches with Digital Mammography for Evaluating Breast Cancer Risk , a Narrative Review," pp. 1110–1119, 2023.
- [16] M. Prodan, E. Paraschiv, and A. Stanciu, "Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection," *Applied Sciences (Switzerland)*, vol. 13, no. 7, 2023, doi: 10.3390/app13074272.
- [17] M. A. P. Wibowo, M. B. Al Fayyadl, Y. Azhar, and Z. Sari, "Classification of Brain Tumors on MRI Images Using Convolutional Neural Network Model EfficientNet," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 538–547, 2022, doi: 10.29207/resti.v6i4.4119.
- [18] Y. Wang, Z. Jin, Y. Tokuda, Y. Naoi, N. Tomiyama, and K. Suzuki, "Development of Deep-learning Segmentation for Breast Cancer in MR Images based on Neural Network Convolution," in *Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition*, New York, NY, USA: ACM, Oct. 2019, pp. 187–191. doi: 10.1145/3373509.3373566.
- [19] K. Jermsittiparsert, A. Abdurrahman, P. Siriattakul, and ..., "Pattern recognition and features selection for speech emotion recognition model using deep learning," *International Journal of ...*, 2020, doi: 10.1007/s10772-020-09690-2.
- [20] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustainability (Switzerland)*, vol. 13, no. 3, 2021, doi: 10.3390/su13031224.
- [21] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Deep Learning-Based Image Segmentation on Multimodal Medical Imaging," *IEEE Trans Radiat Plasma Med Sci*, vol. 3, no. 2, pp. 162–169, Mar. 2019, doi: 10.1109/TRPMS.2018.2890359.
- [22] J. Bar-Ilan, "Which h-index? - A comparison of WoS, Scopus and Google Scholar," *Scientometrics*, vol. 74, no. 2, 2008, doi: 10.1007/s11192-008-0216-y.
- [23] J. Zhang, Q. Yu, F. Zheng, C. Long, Z. Lu, and Z. Duan, "Comparing keywords plus of WOS and author keywords: A case study of patient adherence research," *J Assoc Inf Sci Technol*, vol. 67, no. 4, pp. 967–972, Apr. 2016, doi: 10.1002/asi.23437.
- [24] A. Ninkov, J. R. Frank, and L. A. Maggio, "Bibliometrics: Methods for studying academic publishing," *Perspect Med Educ*, vol. 11, no. 3, 2022, doi: 10.1007/s40037-021-00695-4.
- [25] J. I. Gorraiz, "Editorial: Best Practices in Bibliometrics & Bibliometric Services," *Front Res Metr Anal*, vol. 6, Nov. 2021, doi: 10.3389/frma.2021.771999.
- [26] Ronal Watrianthos, Ambiyar Ambiyar, Fahmi Rizal, Nizwardi Jalinus, and Waskito Waskito, "Research on Vocational Education in Indonesia: A Bibliometric Analysis," *JTEV (Jurnal Teknik Elektro dan Vokasional)*, vol. 8, no. 2, 2022.
- [27] Yose Indarta, R. Watrianthos, and Agariadne Dwinggo Samala, "Research Growth of Engineering Faculty Universitas Negeri Padang: A Bibliometric Study of

- Journal,” *Journal of Systems Engineering and Information Technology (JOSEIT)*, vol. 2, no. 1, pp. 16–24, Mar. 2023, doi: 10.29207/joseit.v2i1.5017.
- [28] C. Birkle, D. A. Pendlebury, J. Schnell, and J. Adams, “Web of science as a data source for research on scientific and scholarly activity,” *Quantitative Science Studies*, vol. 1, no. 1, 2020, doi: 10.1162/qss_a_00018.
- [29] J. Zhu and W. Liu, “A tale of two databases: the use of Web of Science and Scopus in academic papers,” *Scientometrics*, vol. 123, no. 1, 2020, doi: 10.1007/s11192-020-03387-8.
- [30] M. Aria and C. Cuccurullo, “bibliometrix: An R-tool for comprehensive science mapping analysis,” *J Informetr*, vol. 11, no. 4, pp. 959–975, Nov. 2017, doi: 10.1016/j.joi.2017.08.007.
- [31] M. Nagaiah, S. Thanuskodi, and A. Alagu, “Application of Lotka’s Law to the Research Productivity in the field of Open Educational Resources during 2011–2020,” *Library Philosophy and Practice*, vol. 2021, 2021.
- [32] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, “Deep Learning to Improve Breast Cancer Detection on Screening Mammography,” *Sci Rep*, vol. 9, no. 1, p. 12495, Aug. 2019, doi: 10.1038/s41598-019-48995-4.
- [33] T. G. Debelee, S. R. Kebede, F. Schwenker, and Z. M. Shewarega, “Deep Learning in Selected Cancers’ Image Analysis—A Survey,” *J Imaging*, vol. 6, no. 11, p. 121, Nov. 2020, doi: 10.3390/jimaging6110121.
- [34] L. Guo *et al.*, “Divide and Conquer: A Flexible Deep Learning Strategy for Exploring Metabolic Heterogeneity from Mass Spectrometry Imaging Data,” *Anal Chem*, vol. 95, no. 3, pp. 1924–1932, Jan. 2023, doi: 10.1021/acs.analchem.2c04045.
- [35] Z. Sha, L. Hu, and B. D. Rouyendegh, “Deep learning and optimization algorithms for automatic breast cancer detection,” *Int J Imaging Syst Technol*, vol. 30, no. 2, pp. 495–506, Jun. 2020, doi: 10.1002/ima.22400.
- [36] Y. Zhao, J. Zhang, D. Hu, H. Qu, Y. Tian, and X. Cui, “Application of Deep Learning in Histopathology Images of Breast Cancer: A Review,” *Micromachines (Basel)*, vol. 13, no. 12, p. 2197, Dec. 2022, doi: 10.3390/mi13122197.
- [37] Q. Sun *et al.*, “Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don’t Forget the Peritumoral Region,” *Front Oncol*, vol. 10, Jan. 2020, doi: 10.3389/fonc.2020.00053.
- [38] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, “Deep Learning to Improve Breast Cancer Detection on Screening Mammography,” *Sci Rep*, vol. 9, no. 1, p. 12495, Aug. 2019, doi: 10.1038/s41598-019-48995-4.
- [39] F. Jia, X.-C. Tai, and J. Liu, “Nonlocal regularized CNN for image segmentation,” *Inverse Problems & Imaging*, vol. 14, no. 5, pp. 891–911, 2020, doi: 10.3934/ipi.2020041.
- [40] N. A. Farrag, A. Lochbihler, J. A. White, and E. Ukwatta, “Evaluation of fully automated myocardial segmentation techniques in native and contrast-enhanced T1-mapping cardiovascular magnetic resonance images using fully convolutional neural networks,” *Med Phys*, vol. 48, no. 1, pp. 215–226, Jan. 2021, doi: 10.1002/mp.14574.
- [41] L. Zhou, Q. Li, G. Huo, and Y. Zhou, “Image Classification Using Biomimetic Pattern Recognition with Convolutional Neural Networks Features,” *Comput Intell Neurosci*, vol. 2017, pp. 1–12, 2017, doi: 10.1155/2017/3792805.
- [42] J. Alzubi, A. Nayyar, and A. Kumar, “Machine Learning from Theory to Algorithms: An Overview,” *J Phys Conf Ser*, vol. 1142, p. 012012, Nov. 2018, doi: 10.1088/1742-6596/1142/1/012012.
- [43] D. Fareok, R. Umar, and Riadi, “Classification Based on Machine Learning Methods for Identification of Image Matching Achievements,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 198–206, 2022, doi: 10.29207/resti.v6i2.3826.
- [44] L. Sun, S. Zhang, H. Chen, and L. Luo, “Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans With Deep Learning,” *Front Neurosci*, vol. 13, Aug. 2019, doi: 10.3389/fnins.2019.00810.
- [45] H. Seo, L. Brand, L. S. Barco, and H. Wang, “Scaling multi-instance support vector machine to breast cancer detection on the BreaKHis dataset,” *Bioinformatics*, vol. 38, 2022, doi: 10.1093/bioinformatics/btac267.
- [46] E. M. Senan, F. W. Alsaade, M. I. A. Al-Mashhadani, T. H. H. Aldhyani, and M. H. Al-Adhaileh, “Classification of histopathological images for early detection of breast cancer using deep learning,” *Journal of Applied Science and Engineering (Taiwan)*, vol. 24, no. 3, 2021, doi: 10.6180/jase.202106_24(3).0007.
- [47] W. Wiranto and M. R. Uswatunnisa, “Topic Modeling for Support Ticket using Latent Dirichlet Allocation,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 6, pp. 998–1005, 2022, doi: 10.29207/resti.v6i6.4542.
- [48] D. Adimanggala, F. A. Bachtiar, and E. Setiawan, “Evaluasi Topik Tersembunyi Berdasarkan Aspect Extraction menggunakan Pengembangan Latent Dirichlet Allocation,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 511–519, 2021, doi: 10.29207/resti.v5i3.3075.