A Comprehensive Review: Bibliometric Analysis of Decision Tree-Based Approaches for Breast Cancer Prediction

Suhartono*[1], **Syahiduz Zaman** [2]
[1] Universitas Islam Negeri Maulana Malik Ibrahim, Jalan Gajayana 50, malang, Indonesia
suhartono@ti.uin-malang.ac.id
[2] Universitas Islam Negeri Maulana Malik Ibrahim, Jalan Gajayana 50, malang, Indonesia
syahid@ti.uin-malang.ac.id

*Corresponding author

Abstract

This paper aims to conduct a bibliometric analysis of scientific publications that discuss using the decision tree method for breast cancer prediction. Three hundred twenty-two documents from Scopus were collected for analysis using bibliometric indicators such as productivity and citations. The bibliometric analysis produces scientific mapping based on the keywords co-occurrence, co-authorship, and co-citation analysis to reflect the conceptual, social, and intellectual structure. The analysis of the evolution article found an exponential increase in citations and the number of authors in this study in 2005-2023, where China was the dominant country in conducting research. In the thematic map analysis, three research topics were produced, namely the medical field, the computer field, and the bioinformatics field. Research topics in using the decision tree method for breast cancer prediction are included in the computer field. This study suggests that research on using the decision tree method for breast cancer prediction is a research topic that needs improvement.

Keywords: Bibliometric, Decision Tree, Breast Cancer, Prediction

INTRODUCTION

The Decision Tree method is a Machine Learning algorithm for predicting breast cancer in the early stages of diagnosis with a high degree of accuracy, reaching 99% [1]. Previous studies used various machine learning algorithms to build predictive models, such as multilayer perceptron, naive Bayes, and logistic regression [2], [3]. The use of Machine Learning algorithms can improve accuracy in the prediction and diagnosis of breast cancer so that this method can assist in preventing and diagnosing physical injuries [4]. In addition, multiparametric MRI image technology has also been used to classify breast cancer subtypes [5]. The Decision Tree method has been widely used in the prediction and diagnosis of breast cancer with a high degree of accuracy. According to research [6], this method can provide an accuracy of 87.83% with the Gini index and 86.77% with entropy. In addition, using the Decision Tree method with supervised learning techniques can identify the type of cancer more accurately than the Logistic Regression and Neural Network methods [7]. Therefore, the Decision Tree Method has excellent potential to be an effective tool in predicting and diagnosing breast cancer. In a recent study, using the Principal Component Analysis (PCA) method for feature selection in the Wisconsin Diagnostic Breast Cancer dataset can improve performance in the Decision Tree [8]. In addition, a hybrid method is proposed to address the class imbalance problem in breast cancer diagnosis. The method is to combine PCA and Decision Tree with a penalty factor. This method has been tested on breast cancer datasets from the UCI Machine Learning Repository and has proven effective in overcoming this problem [9]. In addition, a study concluded that the Decision Tree method was very effective in detecting breast cancer early, with an accuracy of 100% in the first trial and 97.9% in the second trial [10].

Research results by Assegie [11] showed that the Decision Tree method had an accuracy of 88.80% in predicting breast cancer. In contrast, research conducted by Yang [12] showed that the HER2-IHC staining pattern can be used to predict and diagnose breast cancer. Bibliometric analysis synthesizes research results using mathematical and statistical methods to systematically analyze a literature collection in a particular field [13]. Bibliometric analysis regarding the use of decision trees for breast cancer prediction is still limited. This article aims to collect relevant literature from the Scopus database and conduct a bibliometric analysis to answer three research questions using the Decision Tree method. The first research question is to determine the development trend of applying the Decision Tree method (RQ 1), and the second is to determine the most influential literature, journals, authors, and countries (RQ 2). The last is to find out the characteristics of the literature structure on the Decision Tree (RQ 3).

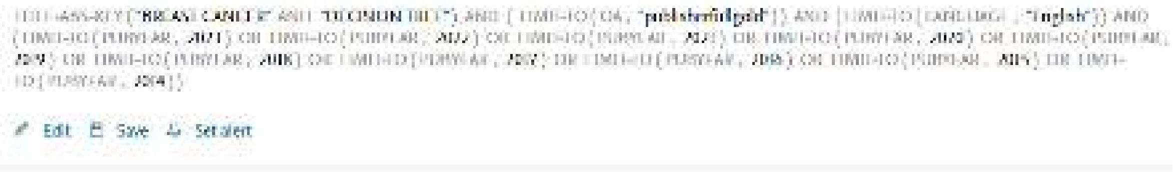
METHODS

Bibliometric analysis is an analysis that uses quantitative methods to analyze a collection of scientific literature. This method analyzes a particular research topic's publications, citations, and joint citations. The bibliometric analysis will discuss the analysis of the number of publications, citations, and joint citations from various articles and journals that discuss this topic. A systematic review was conducted based on the PRISMA guidelines, in which researchers followed statements according to PRISMA 2020 [14]. This study established a systematic review of research to evaluate the use of the Decision Tree method for cancer prediction. By using bibliometric analysis and PRISMA guidelines, this study aims to present evolutionary trends in the application of decision trees for breast cancer prediction, analyze which literature, journals, authors, and countries have the most influence in this research, and identify the characteristics of the structure of the literature regarding the use of decision trees, for breast cancer prediction. This study searched for research articles in the electronic database at Scopus.com to conduct a bibliometric analysis on the use of decision trees in breast cancer prediction. In this study, the Scopus database is used as a data source. One database is used to avoid duplication of literature between databases. The Scopus database is used as a basis for research document search tools and research evaluation. The number of papers Scopus includes reflects the scientific research and capability level. A collection of articles was selected using a query with two aspects: Technique-Related Words (Decision Tree) and Cancer-Related Words (Breast Cancer). This study conducted a bibliometric analysis based on the Scopus database. Some of the analyses carried out were an analysis of the number of papers, an analysis of the number of citations, an analysis of research trends, and an analysis of the most influential journals, authors, countries, and literature in this study, as shown in Table 1.

Table 1. Keywords for research-related

Technique-Related Words	Cancer-Related Words
Decision Tree	Breast Cancer

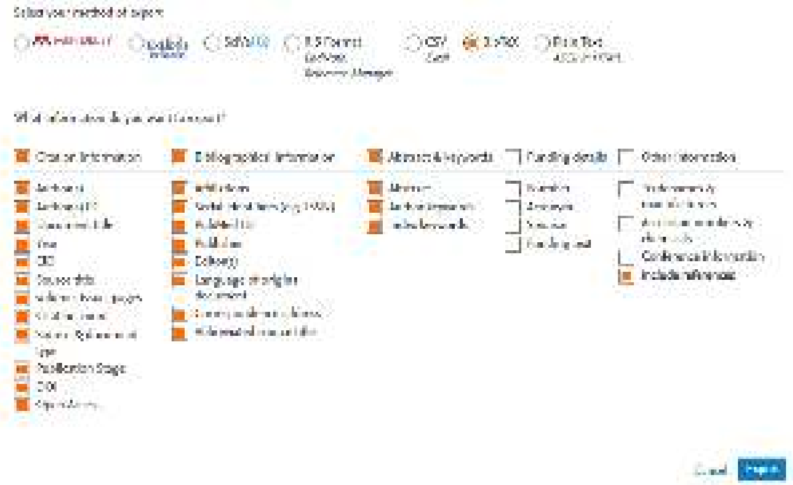
The research comprehensively analyzed the Scopus database by employing the search terms "BREAST CANCER" and "DECISION TREE" to attain the objectives. The search includes title, abstract, author keywords, and keywords in every article in the Scopus database. The search was carried out using specific inclusion and exclusion criteria, such as looking only for articles in English and those published between 2005 and 2023. In this search, we can find articles relevant to the research topic that can be used to carry out bibliometric analysis.



TITLE-ABS-KEY ("BREAST CANCER" AND "DECISION TREE") AND (LIMIT-TO (OA , "publisherfullgold")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014))

Figure 1. Query display on the search tool on the Scopus page

The researcher enters the query according to the research topic, and the next step is to download the data to the bib file. This process makes it easier for researchers to process data and analyze database-based scientific literature. The bib file format is a standard format for collecting references in academic research. Researchers can conduct bibliometric analysis quickly and accurately by collecting references in this format. The download process in the Scopus database can be seen in Figure 2.



Select your method of export

○ Mendeley ○ RefWorks ○ SciVal ○ RIS Format EndNote, Reference Manager ○ CSV Excel ● BibTeX ○ Plain Text ASCII in HTML

What information do you want to export?

☑ Citation information
☑ Author(s)
☑ Author(s) ID
☑ Document title
☑ Year
☑ EID
☑ Source title
☑ volume, issue, pages
☑ Citation count
☑ Source & document type
☑ Publication Stage
☑ DOI
☑ Open Access

☑ Bibliographical information
☑ Affiliations
☑ Serial identifiers (e.g. ISSN)
☑ PubMed ID
☑ Publisher
☑ Editor(s)
☑ Language of original document
☑ Correspondence address
☑ Abbreviated source title

☑ Abstract & keywords
☑ Abstract
☑ Author keywords
☑ Indexed keywords

□ Funding details
□ Number
□ Acronym
□ Sponsor
□ Funding text

□ Other information
□ Tradenames & manufacturers
□ Accession numbers & chemicals
□ Conference information
☑ Include references

Figure 2. Display of downloading data to a bib file via the Scopus page

After obtaining the bib file from the Scopus database, the researcher examined and selected articles according to the research topic using the PRISMA flowchart, as shown in Figure 3. The initial stage is a literature search on the Scopus database, where the query keywords “BREAST CANCER” AND “DECISION TREE” are used to find relevant articles.

Then, an initial selection is made by evaluating the title and abstract, where irrelevant articles are eliminated. After that, further selection was carried out by reading the articles that were still selected, and only articles that met the inclusion and exclusion criteria were selected. The PRISMA process ensured that only relevant and high-quality articles were used in this study for bibliometric analysis.

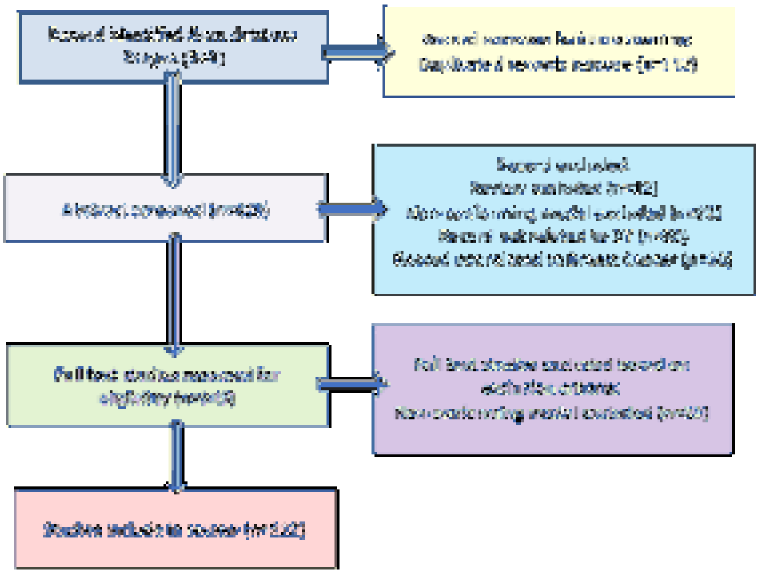


Figure 3. Flow chart for article extraction with PRISMA 2020

In Figure 3, the process with the PRISMA flow resulted in 322 articles, which were selected and included in the bibliometric analysis. The selection of articles covered the period from 2005 to March 2023 and focused on using the Decision Tree method in breast cancer prediction. Some articles that are not relevant to the topic, such as Review excluded (n=12), non-conforming model excluded (n=22), Record not related to DT (n=30), Record not related to Breast Cancer (n=20), were excluded from the analysis. Subsequently, complete articles that met the inclusion criteria were assessed and further analyzed using the bibliometric R package.

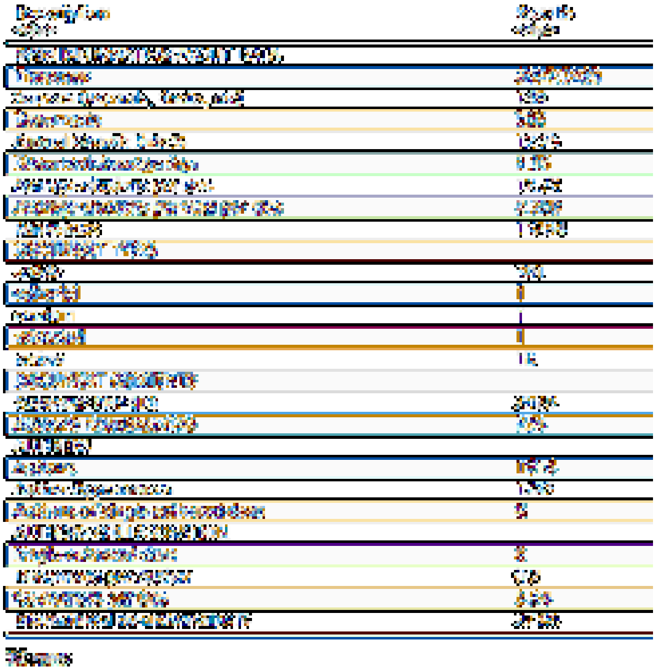
The bibliometric R package is an R language package used as a tool for quantitative research in bibliometrics. Researchers can use this package to analyze bib files systematically. The package provides an open-source environment for bibliometric analysis. The R package bibliometrix (http://www.bibliometrix.org) was employed to conduct trend analysis, descriptive statistical analysis, and author performance analysis on the bib files of the Scopus database.

RESULTS AND DISCUSSION

Descriptive Analysis Of A Bibliographic Data Frame

From the PRISMA process, researchers obtained 322 selected articles. The collection of articles was carried out through descriptive analysis. The results of the descriptive analysis are the primary information collected from the Scopus database for 322 publications published between 2005 and 2023, as shown in Table 2. The selected collection of articles appears in 140 sources, most of which are scientific journal publications with 301 articles. The results of the descriptive analysis stated that the research using the Decision Tree method for predicting breast cancer experienced an increasing trend of 12.25%. “Keywords Plus” is the total number of keywords often appearing in the article’s title. The results of the descriptive analysis state that the number of “Keywords Plus” is 2978. The analysis results can be said to be nine times the number of articles. The research period is from 2005 to 2023 (17 years), and the number of publications increases by 12.25% yearly, as shown in table 2.

Table 2. Results of descriptive analysis



[illegible]	[illegible]
[illegible]	
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

[illegible]

Analysis Of Performance

To answer RQ1 and RQ2, we analyze the performance of articles using several bibliometric indicators such as number of publications, number of citations, h-index, m-quotient, and several variable metrics based on the number of publications and citations. The analysis results concluded that publications about using the Decision Tree method for breast cancer prediction have increased significantly in the last 17 years, with the number of publications reaching 322. The analysis results also produce output about the most influential authors, sources, and countries in this research, such as authors who are often cited, journals that are often published, and countries that are active in this research.

Analysis of Evolution Article

Three hundred twenty-two articles relevant to the research topic have been found from 2005 to 2023. In Figure 4, we can see that the number of publications has increased significantly every year. The highest peak will be in 2022 when the number of publications will be 99 articles. The number of publications was stagnant at two from 2005 to 2013. There was an increase in publications from 2014 to 2022. This increase shows that the Decision Tree method is increasingly in demand in the context of breast cancer prediction. More and more researchers are interested in developing and improving this technique. It becomes the basis for researchers and practitioners in developing new and up-to-date strategies in breast cancer prediction using the Decision Tree method.

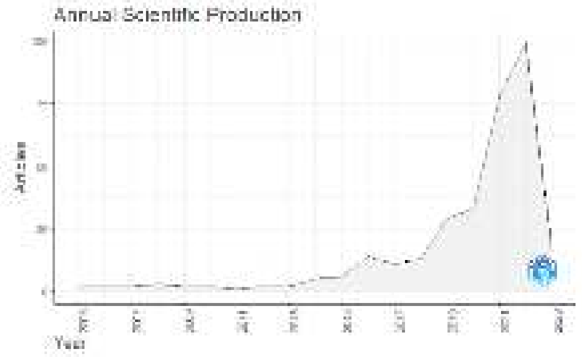


Figure 4. Distribution of the number of publications each year (2005-2023)

Whereas citations in research using the Decision Tree method for predicting breast cancer showed stagnant growth from 2005 to 2023, as shown in Figure 5, the increase in citations occurred twice, namely in 2011 and 2019.

The number of citations each year does not show a particular trend. The maximum number of citations occurred in 2011 and 2019, and the maximum number of citations was 6, while the minimum number of citations was in 2007 and 2014, and the minimum number of citations was 1, as shown in Figure 5.

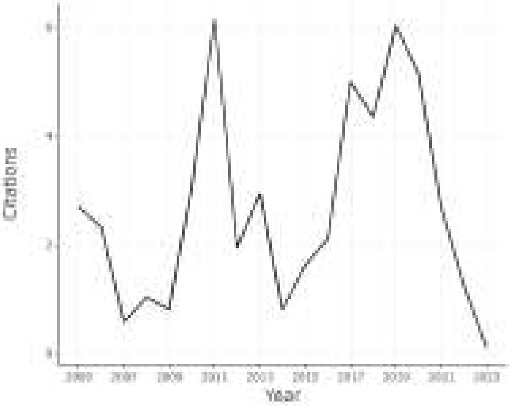
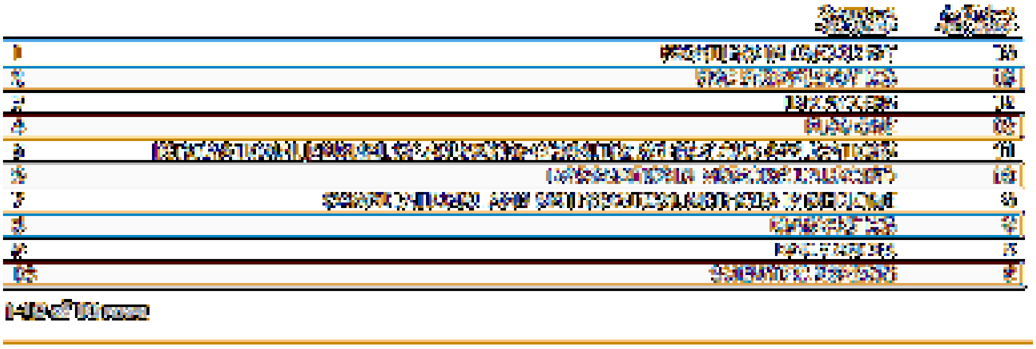


Figure 5. Distribution of citations every year from 2005 to 2023

Analysis of Important Journal

In this bibliometric analysis, there are 322 articles published in 140 journals. Table 3 shows the top 10 journals regarding the number of publications. The ranking of journals in the table is based on the number of publications (TP). The journal that ranks first is “Frontiers in Oncology” with 19 published articles, followed by “BMC Bioinformatics” with 13 articles, and “IEEE Access” with 12 articles. The fourth and fifth positions are “Plos One” and “International Journal of Advanced Computer Science and Applications,” with 12 published articles. Other journals included in the top 10 list are “Informatics in Medicine Unlocked," “Computational and Mathematical Methods in Medicine," “Diagnostics," “BMC Cancer,” and “Scientific Report”. The ranking of these journals can be used as a reference in determining publications in related journals for further research in the field of using decision trees.

Table 3. Top 10 journals in terms of number of publications



	[illegible]	[illegible]
1	[illegible]	[illegible]
2	[illegible]	[illegible]
3	[illegible]	[illegible]
4	[illegible]	[illegible]
5	[illegible]	[illegible]
6	[illegible]	[illegible]
7	[illegible]	[illegible]
8	[illegible]	[illegible]
9	[illegible]	[illegible]
10	[illegible]	[illegible]

[illegible]

Analysis of Citations Number

Analysis of the number of citations received by an article is one way to find influential articles in their field. In the study of breast cancer prediction using the Decision Tree method, many articles have been published from 2005 to 2023. Table 4 contains this study's ten most frequently cited articles, sorted by the number of citations received. Based on Table 4, the title of the most cited article is “Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms” by author Pelon [15]. The number of citations is 163 citations. This article contributes significantly to deepening the understanding of using the Decision Tree in breast cancer prediction. Also, it provides insight into the role of CAF in breast cancer metastasis. In addition, the articles included in the ten most frequently cited articles, besides discussing the decision tree method, also discuss various techniques such as ensemble learning, Naïve Bayes, and Principal Component Analysis. The article title with the second highest number of citations is “Ensemble of decision tree reveals potential miRNA-disease associations” by author Chen [8], with 131 citations. The article discusses a new method called the Ensemble of Decision Tree-based miRNA-disease Association Prediction (EDTMDA) to predict potential miRNA disease associations.

Table 4. The ten most cited articles

Author	Title	Citations
[15]	Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms	163
[8]	Ensemble of decision tree reveals potential miRNA-disease associations	131
[16]	Predicting factors for survival of breast cancer patients using machine learning techniques	110
[17]	Machine learning with applications in breast cancer diagnosis and prognosis	107
[18]	Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction	104
[19]	A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning	99
[20]	Validation of cytoplasmic-to-nuclear ratio of survivin as an indicator of improved prognosis in breast cancer	80
[21]	Involvement of Machine Learning for Breast Cancer Image Classification: A Survey	73
[22]	A new fruit fly optimization algorithm enhanced support	65

	vector machine for diagnosis of breast cancer based on high-level features	
[23]	Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification	61

Analysis of Significant Author

Data in unstructured text, such as articles or documents, has limited or no structure. This form causes computer problems with working on semantics. However, with bibliometric analysis, the text data can be analyzed by extracting the metadata that represents its features. These features can be converted into a structured form, an intermediate document representation. Based on Table 5, author Ganggayah [16] is the first highest collaborative writer. Author Kaur produced one article relevant to the topic and received 132 citations. This study proposed a new approach for breast cancer diagnosis and prediction using Deep Learning (DL) and Support Vector Machine (SVM) methods with an average accuracy of 95%, 94%, and 98% for three classes and higher sensitivity and specificity than other methods. Chen H is a collaborative researcher. Chen H is the collaborative author with the second-highest number of articles cited; author Chen H produced four relevant articles. He proposed Machine learning with applications in breast cancer diagnosis and prognosis and received more than 107 citations. Research discussing breast cancer and early diagnosis and accurate classification of patients into malignant or benign groups is fundamental [17].

Table 5. Ranking of the most productive authors

Author	year	freq	TC	TCpY	Author
Chen H	2018	1	107	17,833	[17]
	2019	1	65	13	[22]
	2021	1	0	0	[24]
	2022	1	0	0	[25]
Fusco R	2017	1	24	3,429	[26]
	2021	2	8	2,667	[27], [28]
	2023	1	0	0	[29]
Kaur P	2019	1	132	26,4	[23]
Li H	2018	1	6	1	[12]
	2019	1	16	3,2	[30]
	2021	1	16	5,333	[31]
	2022	1	0	0	[32]
Li J	2018	1	6	1	[12]
	2019	1	16	3,2	[30]
	2021	2	46	15,333	[33],[34]
	2022	1	0	0	[35]
Li L	2021	1	3	1	[36]
	2022	2	2	1	[32],[37]
Li X	2018	1	6	1	[12]
	2019	1	16	3,2	[30]
	2022	2	1	0,5	[38],[39]
Liu Y	2006	1	53	2,944	

	2018	1	6	1	[12]
	2021	1	3	1	[36]
	2022	2	0	0	[40], [41]
Wang X	2021	2	9	3	[42], [43]
	2022	2	0	0	[39],[25]
Zhang Y	2018	1	15	2,5	[44]
	2019	1	16	3,2	[30]
	2020	2	16	4	[45],[46]
	2022	3	3	1,5	[4],[47],[41]

In the bibliometric analysis of this study, articles were still found with a low number of citations (0 citations). According to the researchers, the main reasons for the low number of citations were a lack of understanding of collaborative research at home and abroad, lack of innovation, or ignoring domestic peer references. The author also divides one article into two or more to increase the number of publications for the author, which can affect the quality of the paper. The annual distribution of the number of publications is plotted in Figure 6 for the author's performance. Of the ten authors with the most publications, the research trend continues to increase, as shown by nine authors still writing articles. The size of the circle on the chart indicates the number of papers published by the author, and the bigger the circle, the more papers published in that year. In addition, the circle’s color represents the number of times the author has been cited, and the darker the color of the circle, the higher the number of citations received by the author that year. Author Kaur P published 1 article in 2019 with 132 citations per year, the smallest number of publications but the highest number of citations.

Author Liu Y was the first to introduce this method for breast cancer prediction in 2006. Liu Y published an article in Cancer Informatics discussing a hybrid approach to discovering cancer biomarkers from gene expression data [48]. Liu Y also compared six filter methods and three wrap methods to determine biomarkers and developed a hybrid approach that combines the two methods. Author Fusco became the second pioneer to publish an article on breast cancer prediction using dynamic and morphological features with several classifications [26].

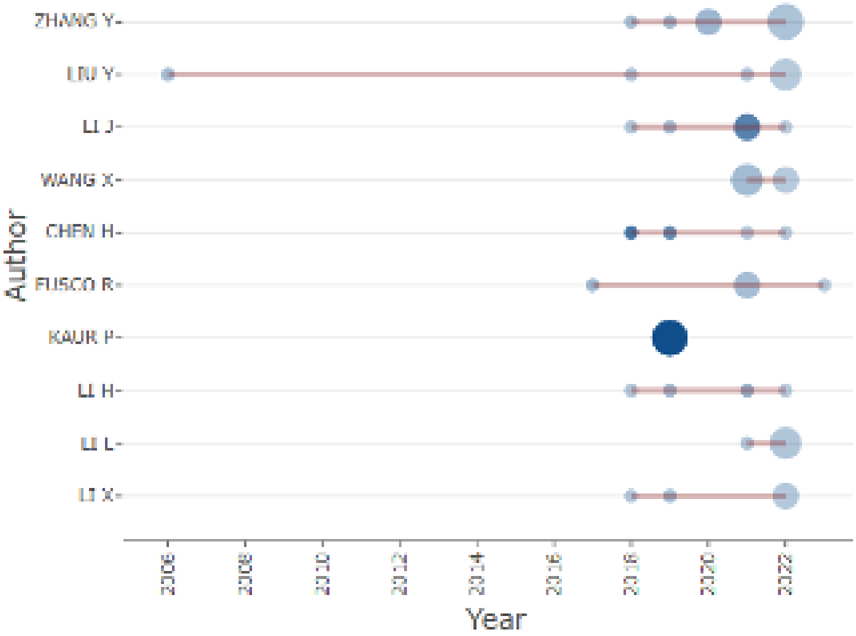


Figure 6. Number of articles in the top 10 authors over time.

Analysis of Science Mapping

Scientific mapping analysis explains how the conceptual structure, co-occurrence network, and intellectual literature are characterized. Scientific mapping will reveal information about the topic groups discussed, trending articles often cited, and keywords often used. In addition, this analysis will evaluate collaborations between authors and identify the most influential authors and journals.

Analysis of Conceptual Structure

An article is a scientific report; articles are essential in a scientific information guide. Keywords in articles can help in searching and saving an article. In addition, Keyword Plus can be used to increase search articles based on keywords or titles. To form a Keyword Plus co-occurrence network, you can use the ‘biblioNetwork’ function in the ‘bibliometrix’ package. In Figure 7, the number and ranking of keywords frequently appearing in articles can provide an overview of the research topic. Therefore, Keyword and Keyword Plus are very important in bibliometric analysis and can assist researchers in exploring a collection of articles related to a particular topic. The keywords Breast Cancer and Decision Tree are close together, so it can be said that they are often used together in explanations in articles. The keyword decision tree ranks 6th, is often used in a collection of articles, and has been used for research since 2016.

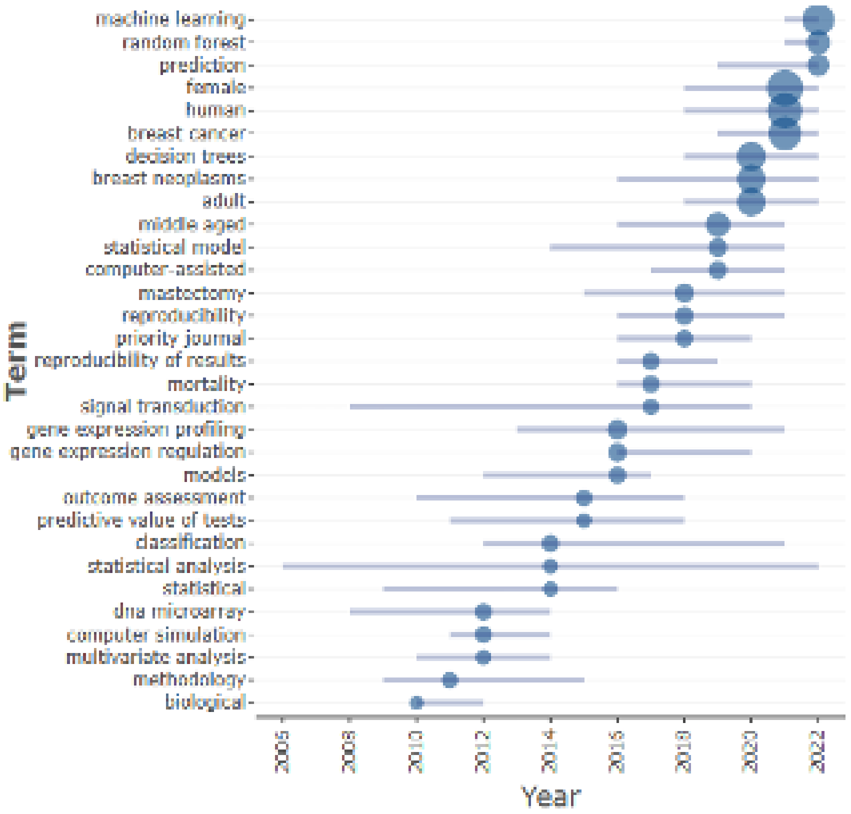


Figure 7. Keywords in articles that often appear

Analysis of co-occurrence network

In the co-occurrence network analysis of the plus keyword, the network has three main groups in this collection of articles, as shown in Figure 8. The first group consists of keywords related to breast cancer concepts in the medical field, such as “breast cancer," “mammary carcinoma,” and “breast neoplasm.” This group is marked with a green circle. While the second group consists of keywords related to breast cancer analysis methods in the computer field, such as “decision tree," “machine learning,” and “statistical analysis.” This group is marked with a red circle. In addition, several keywords appear separately from the two main groups, such as “mammary carcinoma” and “breast neoplasm," “predictive model," “diagnostic,” and “genetic algorithm.” This group is related to the concept of breast cancer in bioinformatics. This group is marked with a blue circle.

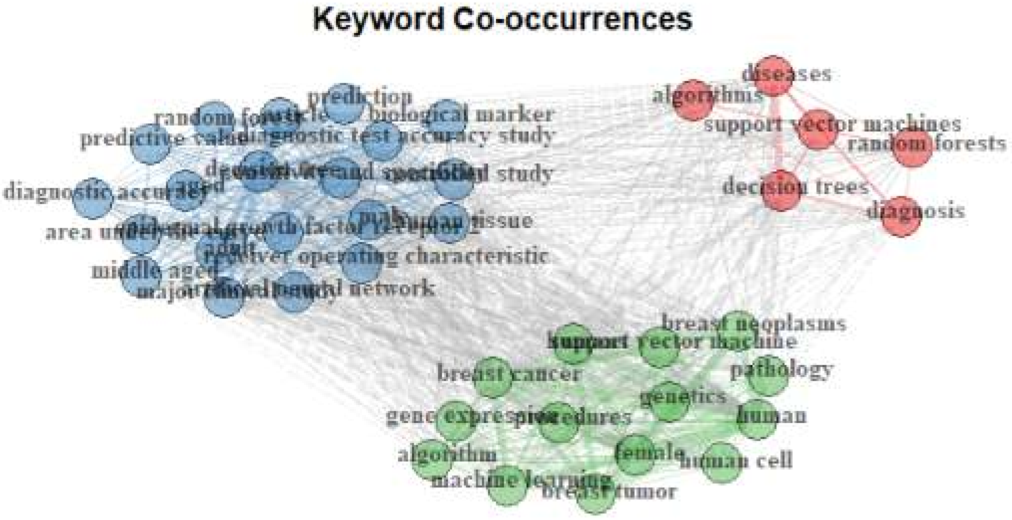


Figure 8. Co-occurrence network of plus keywords

Analysis of conceptual structure map

Conceptual structure map analysis can be employed to facilitate identifying articles relevant to research topics. In this conceptual structure map, 250 keywords appear in articles. From 734 keywords and 2974 keywords, 250 keywords were taken. The 250 keywords were categorized into three research theme groups. The first group is the theme of breast cancer research in the medical field. This group is marked with a green circle. At the same time, the second group is the theme of breast cancer research in medicine. This group is marked with a red circle. The third group is the theme of breast cancer research in the field of bioinformatics. This group is marked with a blue circle.

The grouping results are then presented as a strategic diagram in Figure 9. This diagram can be used to visualize the relationship between keywords and research themes, making it easier for users to find articles related to research topics. Figure 9 explains the strategic diagram, which consists of three research topics based on a collection of articles. Each research topic in the strategic diagram has two parameters: Centrality and Density. The horizontal axis measures centrality and strengthens the external connection between the research topic and the research topic in collecting articles. This parameter measures the importance of research topics in research development. The vertical axis measures density. This parameter is the strength of the connection between keywords in the research topic. Density can be used to measure the degree of the research topic. This map shows that the theme of breast cancer research in the computer field has low Centrality and Density. In contrast, the theme of breast cancer research in the medical field has high density, and the theme of breast cancer research in the bioinformatics field has high centrality.

The strategic diagram in Figure 9 can also provide information about the position of each research theme in the development of the breast cancer research field. The upper right quadrant of the strategic diagram shows that the themes related to breast cancer research have high density and centrality, so the themes are well-developed and essential for constructing research fields. The lower right quadrant displays research themes with high centrality but low density, indicating that these are important for the development of the research field but are not well developed and are generally the primary themes in the research field. Themes with high density but low centrality are found in the upper left quadrant, indicating that the research theme is well-developed but has a limited impact on the research field. At the same time, the lower left quadrant highlights themes with low density and centrality, which can be considered as less essential themes in the field of research. The thematic map analysis in Figure 9 in this study has three main research topics. Even though the topic of breast cancer research in the computer field shows low density and low centrality, this research topic is still considered essential to be developed because it has the potential to become a center for the development of all breast cancer research. Second, although the number of research articles related to the computer field is still tiny compared to the medical and informatics fields, based on thematic map analysis, the computer field is still developing and requires attention.

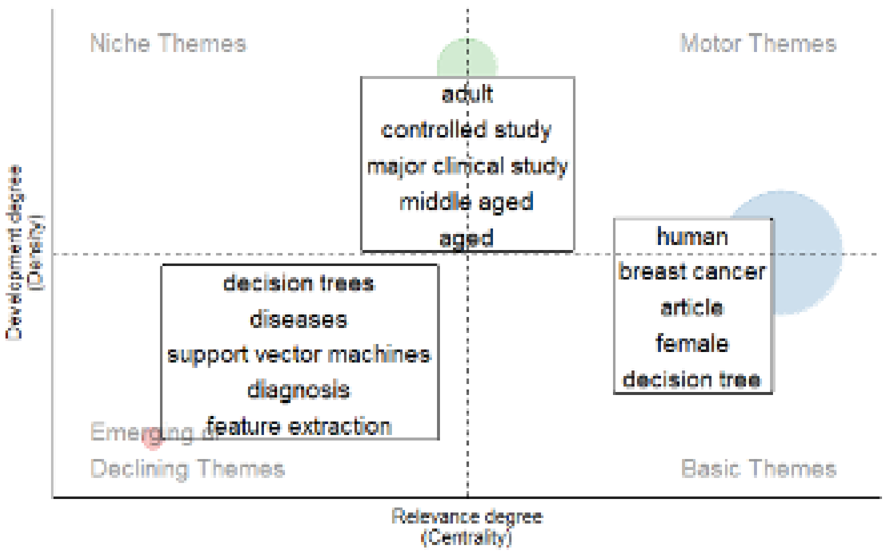


Figure 9. Thematic map based on keywords

Analysis of the most contributed articles

The following are the fifteen articles that have contributed the most to each research topic (five articles); as shown in Table 7, there are five articles with high contributions to the research topic in the field of medicine (green color), namely Pelon [15], Ferraro [49], Rexhepaj [20], Hadi [50] and Sánchez-Calderón [51]. The article by Pelon [15] used cancer-associated fibroblast heterogeneity in the axillary lymph nodes. In contrast, the article by Ferraro [49] used a microfluidic platform that combines a liquid drop and magnetic tweezers. The article by Rexhepaj [20] used the cytoplasm-to-nucleus ratio of confirmed survival as a prognostic indicator. In contrast, the article by Hadi [50] used a Serum Metabolomics profile for Diagnosis, Classification, and Stage of Breast Cancer. Meanwhile, the article by Sánchez-Calderón [51] used a cost-benefit analysis of liquid biopsies to determine patient treatment changes. The other five articles that best represent research themes in the computer field (red color) are Sakri [18], Abbas [52], Saarela [53], Osman [54], and Kaya Keleş [55]. Author Sakri [18] used particle swarm optimization for feature selection in predicting breast cancer recurrence, while the article by Abbas [52] used the BCD-WERT approach for breast cancer detection using efficient features. The article by Saarela [53] uses a comparison of feature importance measures as an explanation for the classification model. In contrast, the article by Osman [54] uses the Ensemble Boosting Learning Method, which is effectively used for virtual breast cancer screening using the Neural Network model. Meanwhile, author Kaya Keleş [55] uses a data mining classification algorithm to predict and detect breast cancer.

The other four articles that most represent the field of bioinformatics (blue color) are Chen [8], Chowdhary [56], Jayatilake [57], and Al-Azzam [58]. In this research, researchers continue to develop bioinformatics approaches to improve the diagnosis and treatment of diseases, especially breast cancer. Author Chen [8] used an Ensemble from a decision tree to uncover potential miRNA-disease associations. In contrast, Chowdhary [56] used an efficient segmentation and classification system for medical images using intuitionist-possibilistic fuzzy C-mean clustering and the fuzzy SVM algorithm. In addition, Jayatilake [57] utilizes machine learning tools to assist in health decision-making, while Al-Azzam [58] compares Supervised and Semi-Supervised Machine Learning Models in Diagnosing Breast Cancer. To obtain more information about the contribution of papers related to the three topics in the research field, Table 7 shows the five articles with the most considerable contributions for each topic. The first topic is related to the field of computers; a paper with a significant contribution is "A Novel Hybrid Feature Selection Approach Using a Decision Tree and Artificial Bee Colony Algorithm for Breast Cancer Diagnosis" by Keles [55]. The second topic is related to bioinformatics; the paper with the most significant contribution is "Artificial Intelligence for Breast Cancer Detection in Mammography: A Systematic Review" by Al-Azzam [58].

Table 7. The articles that most contributed to the research topic

Year	Contribution	Citations
		Author
2020	40,75	[15]
2016	5,625	[49]
2010	2,929	[20]
2017	5,714	[50]
2020	3,25	[51]
2018	17,333	[18]
2021	18,667	[52]
2021	18,333	[53]
2020	7,25	[54]
2019	5,4	[55]
2019	26,2	[8]
2020	12,75	[56]
2021	14,667	[57]
2021	11,333	[58]

Analysis of the Social Order

Grouping of author collaborations uses the Louvain Algorithm. The results of the Louvain Algorithm are in the form of visualization; visualization can explain the relationship between authors in collaborative article writing, as shown in Figure 10. In visualization, 30 of the most influential writers collaborated on research according to the topic during 2005-2023. This visualization provides insight into research collaboration among authors and helps understand scientific networks.

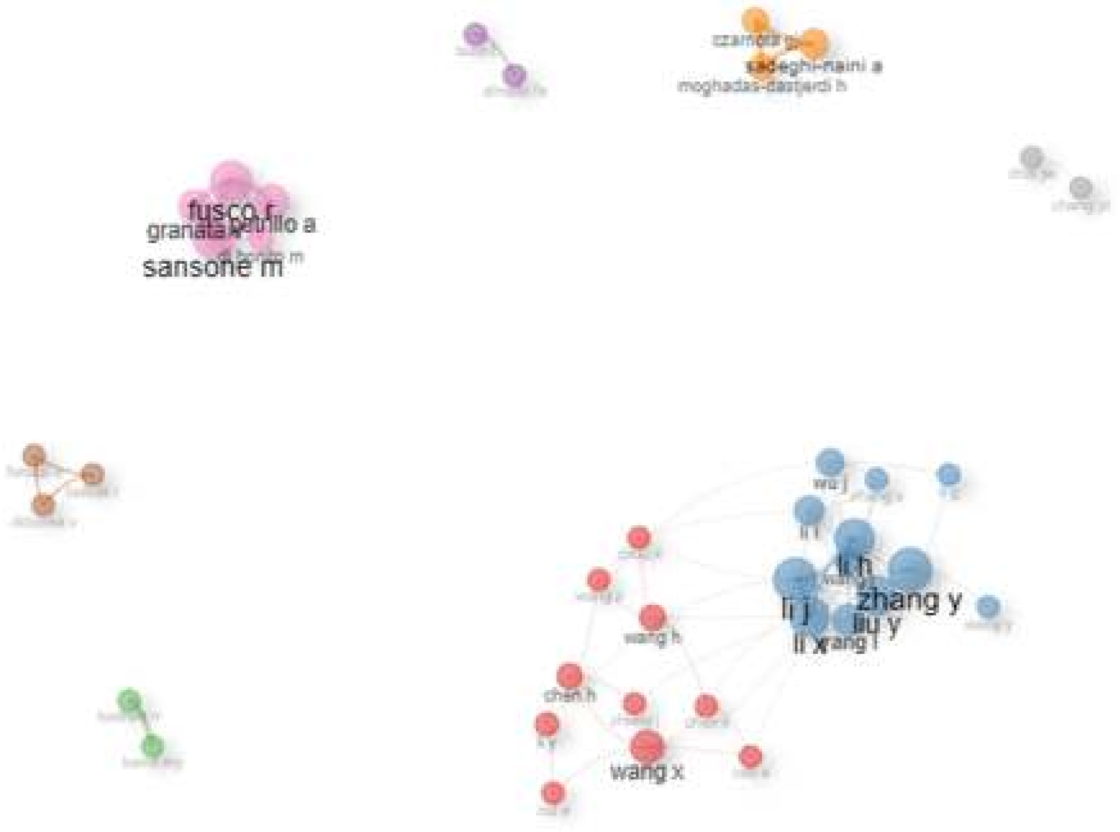


Figure 10. Thematic map based on keywords

The visual representation of author collaboration through co-occurrence technique is shown in Figure 10. The size of the circles indicates the number of articles written by the authors, and the cross coverage between the circles indicates the number of articles written together by the authors. It should be noted that isolated nodes do not necessarily mean that the authors did not collaborate with others but only prove that they did not collaborate with the top 30 influential authors. Therefore, these isolated nodes are removed from the figure to avoid misinterpretation. The figure shows eight collaborative groups, divided into three categories, are among the top 30 influential authors. The collaborative group led by Liu Y has the most frequent collaborative relationship, as indicated by the dark blue color. Visualization explains the author's collaboration through the co-occurrence technique, as shown in Figure 8. The circle size indicates the number of articles co-authored by the authors, and the cross coverage between circles indicates the number of articles co-authored by the authors.

It should be noted that isolated nodes do not mean that the author did not collaborate with others but only prove that they did not collaborate with the top 30 influential authors. As seen in Figure 8, there are eight collaborative groups among the top 30 influential authors. The collaborative group led by Liu Y has the most frequent collaborative relationships shown in blue. Figure 9 shows the country of origin of the authors of articles on using the Decision Tree for breast cancer prediction worldwide. The big circle represents the country represented by the number of national correspondences, while the thickness of the relationship between countries shows the frequency of collaboration between these countries.

It can be seen that the United States and China are the countries with the darkest color, indicating that researchers from these countries have written the highest number of articles, and the results of the analysis are consistent. The most frequent link was between the US and China, indicating that researchers from the two countries collaborated and communicated more frequently. In addition, India and Iran also occupy the second position in the frequency of collaboration, as shown in Figure 11.

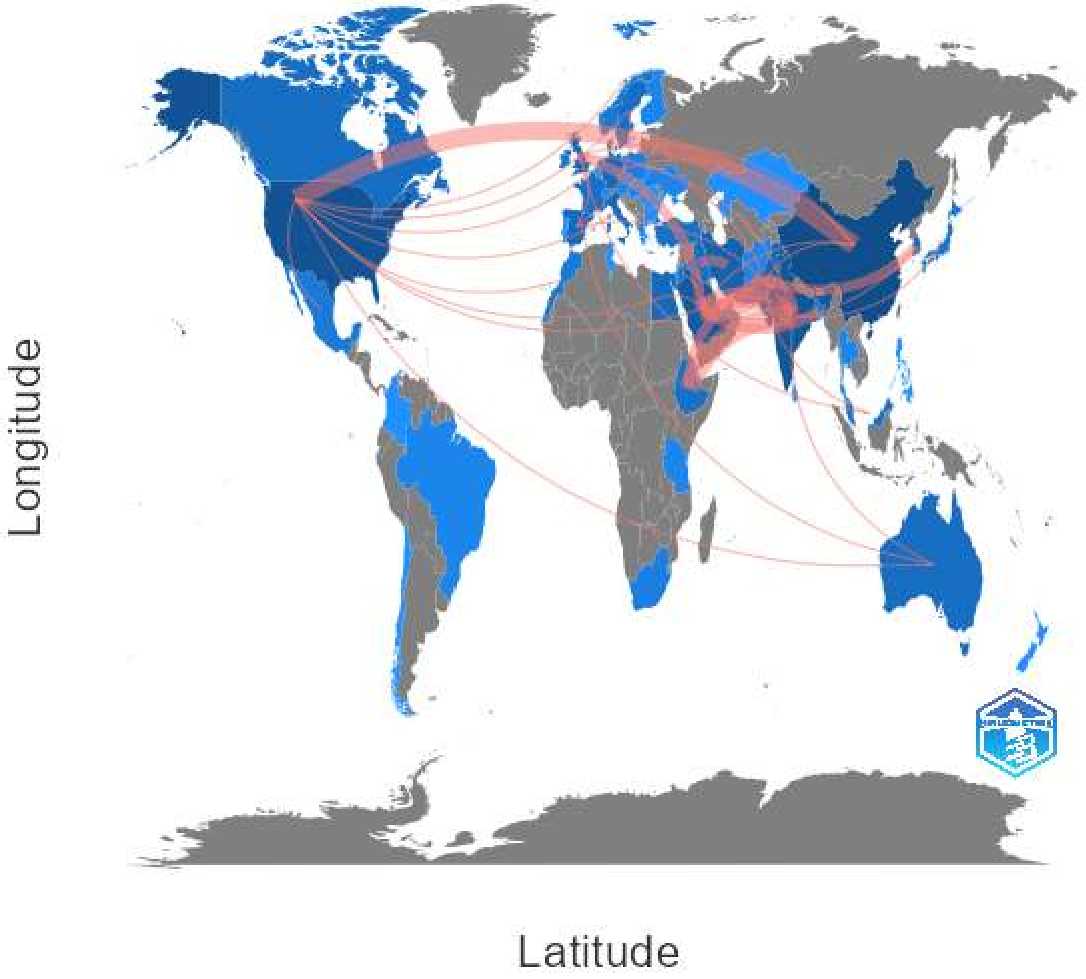


Figure 11. Level of collaboration between countries based on author affiliation.

Analysis of Struktur Intelektual

Figure 12 displays research historiography using the Decision Tree method for breast cancer prediction based on citation networks. In the 20 articles analyzed, starting from the articles written by [16] until the article, there are three citation networks with different colors: blue, red, and green. The first cluster in red font with eight authors consists of authors Shanbehzadeh [59], Li [34], Massafra [60], Gangayah [16], Qawqzeh [61], Nik Ab Kadir [62], and Ozcan [63]. The second cluster in red font with two authors consists of Abbas [52] and Zhang [30]. The third cluster is in green font with three authors: Osman [63], Huang [64], and Khan [65]. Figure 12 shows the citation relationship between the authors involved in this study during 2019-2023.

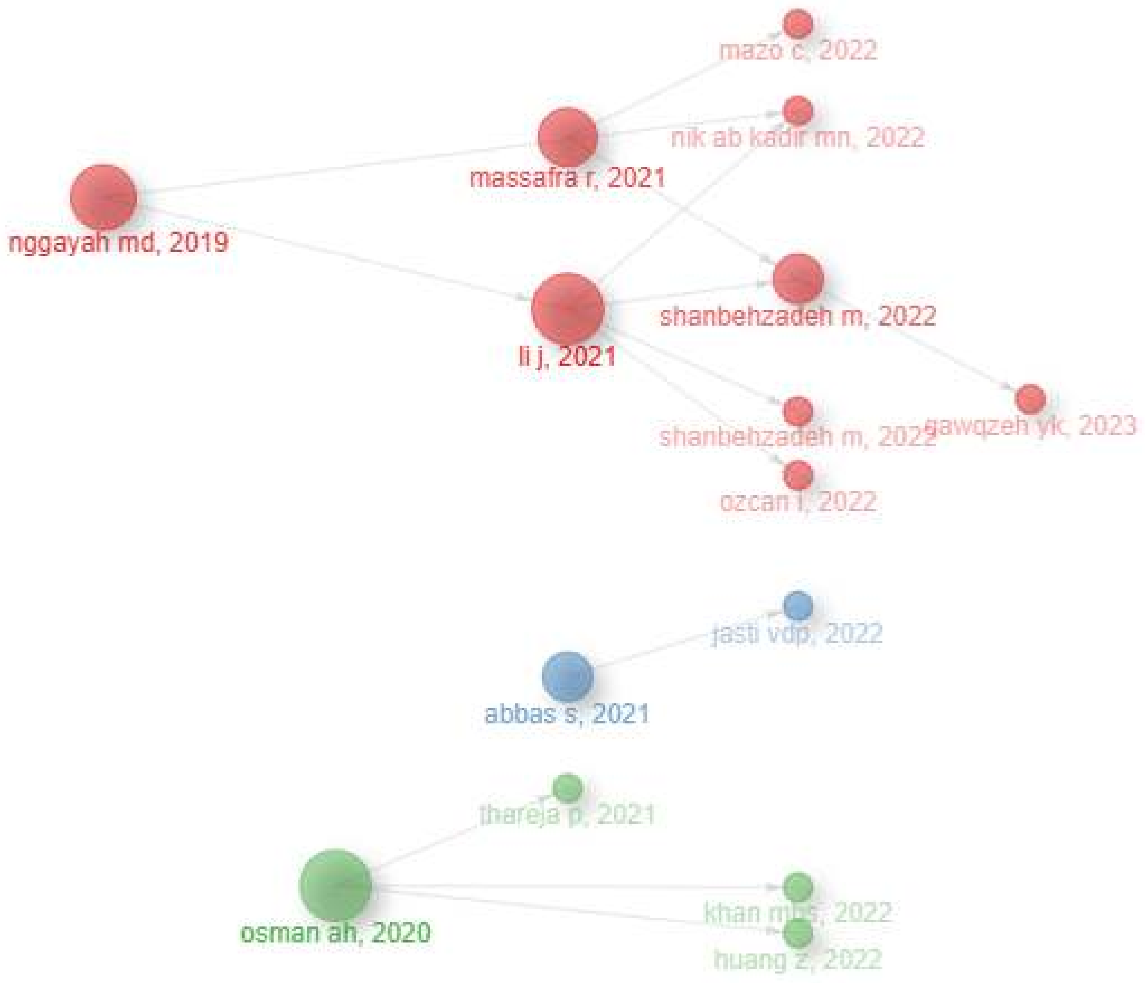
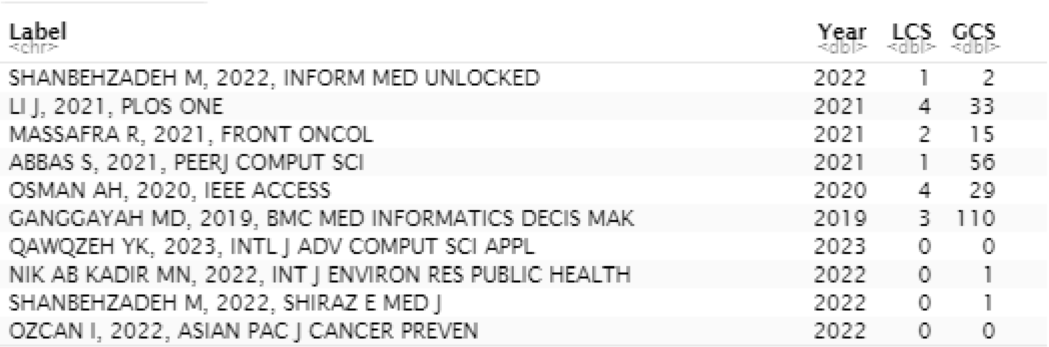


Figure 12. Historical evolution of shared citations among the 20 most relevant articles.

Figure 12 shows the historical development of 20 influential articles on breast cancer prediction according to chronological and cluster order. In the first cluster, there is the first most influential author, Gangyah [16], with the article "Predicting factors for survival of breast cancer patients using machine learning techniques." This article has a local citation score (LCS) of 3 and a global citation score (GCS) of 110, published by BMC Medical Informatics and Decision Making in 2019, as shown in Table 8. This study built a predictive model using a decision tree, random forest, neural networks, extreme boost, logistic regression, and a support vector machine. Important variables identified included cancer stage classification, tumor size, total number of axillary lymph nodes removed, number of positive lymph nodes, type of primary treatment, and method of diagnosis. All algorithms produce similar accuracy, and machine learning methods can be alternative predictive tools[16].

Table 8. The information on 20 articles for the citation network.



Label	Year	LCS	GCS
<chr>	<dbl>	<dbl>	<dbl>
SHANBEHZADEH M, 2022, INFORM MED UNLOCKED	2022	1	2
LI J, 2021, PLOS ONE	2021	4	33
MASSAFRA R, 2021, FRONT ONCOL	2021	2	15
ABBAS S, 2021, PEERJ COMPUT SCI	2021	1	56
OSMAN AH, 2020, IEEE ACCESS	2020	4	29
GANGGAYAH MD, 2019, BMC MED INFORMATICS DECIS MAK	2019	3	110
QAWQZEH YK, 2023, INTL J ADV COMPUT SCI APPL	2023	0	0
NIK AB KADIR MN, 2022, INT J ENVIRON RES PUBLIC HEALTH	2022	0	1
SHANBEHZADEH M, 2022, SHIRAZ E MED J	2022	0	1
OZCAN I, 2022, ASIAN PAC J CANCER PREVEN	2022	0	0

Articles written by Ganggayah [16] have become the basis for citations of other articles in subsequent years, including an article by Li J in 2021[34]. This article describes past research using Machine Learning (ML) to predict patient survival rates for breast cancer for five years. It was found that 31 articles met the inclusion criteria; most were published after 2013 and used the most frequently used ML methods, such as Decision Trees, Artificial Neural Networks, Vector Support Machines, and Ensemble Learning. However, the results of the analysis show that the ML model's performance has not significantly improved compared to traditional statistical methods [34].

A study conducted by Ganggayah [16] regarding using machine learning techniques to build a predictive model for breast cancer received quotes from several recent articles. One of them is a study by Li J in 2021 [34] regarding a systematic review of research using Machine Learning to predict the survival rate of breast cancer patients for five years. In this research, Decision Trees are the most frequently used Machine Learning method with model accuracy that has not shown significant improvement compared to traditional statistical methods. Then, in 2022, Nik Ab Kadir MN [62] researched Malaysia to develop a predictive model for survival among women with breast cancer using Cox proportional hazards, artificial neural networks, and classification decision tree analysis. In that study, the Cox PH model had the highest accuracy in predicting patient survival compared to the DT and ANN models [62].

CONCLUSIONS

Bibliometric analysis can be used to analyze author performance and scientific mapping in the structure of scientific publications. Author performance analysis evaluates articles' productivity and impact using bibliometric indicators. Likewise, scientific mapping shows conceptual, social, and citation structures using co-occurrence analysis on keyword plus, co-author, and co-reference analysis. To analyze bibliometrics, use the Bibliometric in R package. They were using the Scopus database for literature collections. As a limitation, keywords and years were used to limit the literature analyzed. This research found that the three countries that produced the most articles were China, USA, and Iran. In comparison, the three countries that produced the most citations were China, the USA, and India. Further studies can be carried out by adding to the literature collection and being selective about the keywords used to get more comprehensive findings.

REFERENCES

[1] A. El-Nabawy, N. A. Belal, and N. El-Bendary, "A cascade deep forest model for breast cancer subtype classification using multi-omics data", *Mathematics*, vol. 9, no. 13, 2021, doi: 10.3390/math9131574.

[2] M. M. Alshammari, A. Almuhanna, and J. Alhiyafi, "Mammography image-based diagnosis of breast cancer using machine learning: A pilot study", *Sensors*, vol. 22, no. 1, 2022, doi: 10.3390/s22010203.

[3] M. Botlagunta *et al.*, “Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms”, *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-27548-w.

[4] G. Li, T. Fang, Y. Zhang, C. Liang, Q. Xiao, and J. Luo, “Predicting miRNA-disease associations based on graph attention network with multi-source information”, *BMC Bioinformatics*, vol. 23, no. 1, 2022, doi: 10.1186/s12859-022-04796-7.

[5] T. Xie *et al.*, “Machine learning-based analysis of MR multiparametric radiomics for the subtype classification of breast cancer”, *Front. Oncol.*, vol. 9, no. JUN, 2019, doi: 10.3389/fonc.2019.00505.

[6] F. K. Nasser and S. F. Behadili, “Breast Cancer Detection using Decision Tree and K-Nearest Neighbour Classifiers”, *Iraqi J. Sci.*, vol. 63, no. 11, pp. 4987–5003, 2022, doi: 10.24996/ijs.2022.63.11.34.

[7] V. R. Mudunuru and L. A. Skrzypek, “A comparison of artificial neural network and decision trees with logistic regression as classification models for breast cancer survival”, *Int. J. Math. Eng. Manag. Sci.*, vol. 5, no. 6, pp. 1170–1190, 2020, doi: 10.33889/IJMEMS.2020.5.6.089.

[8] X. Chen, C.-C. Zhu, and J. Yin, “Ensemble of decision tree reveals potential miRNA-disease associations”, *PLoS Comput. Biol.*, vol. 15, no. 7, 2019, doi: 10.1371/journal.pcbi.1007209.

[9] J.-X. Tian and J. Zhang, “Breast cancer diagnosis using feature extraction and boosted C5.0 decision tree algorithm with penalty factor”, *Math. Biosci. Eng.*, vol. 19, no. 3, pp. 2193–2205, 2022, doi: 10.3934/MBE.2022102.

[10] L. Dobrovska and O. Nosovets, “Development Of The Classifier Based On A Multilayer Perceptron Using Genetic Algorithm And Cart Decision Tree”, *East.-Eur. J. Enterp. Technol.*, vol. 5, no. 9–113, pp. 82–90, 2021, doi: 10.15587/1729-4061.2021.242795.

[11] T. A. Assegie, R. L. Tulasi, and N. K. Kumar, “Breast cancer prediction model with decision tree and adaptive boosting”, *IAES Int. J. Artif. Intell.*, vol. 10, no. 1, pp. 184–190, 2021, doi: 10.11591/ijai.v10.i1.pp184-190.

[12] L. Yang *et al.*, “A decision tree-based prediction model for fluorescence in situ hybridization HER2 gene status in HER2 immunohistochemistry-2+ breast cancers: A 2538-case multicenter study on consecutive surgical specimens”, *J. Cancer*, vol. 9, no. 13, pp. 2327–2333, 2018, doi: 10.7150/jca.25586.

[13] 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.

[14] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”, *Syst. Rev.*, vol. 10, no. 1, p. 89, Dec. 2021, doi: 10.1186/s13643-021-01626-4.

[15] F. Pelon *et al.*, “Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms”, *Nat. Commun.*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-019-14134-w.

[16] M. D. Ganggayah, N. A. Taib, Y. C. Har, P. Lio, and S. K. Dhillon, “Predicting factors for survival of breast cancer patients using machine learning techniques”, *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 48, Dec. 2019, doi: 10.1186/s12911-019-0801-4.

[17] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, “Machine learning with applications in breast cancer diagnosis and prognosis”, *Designs*, vol. 2, no. 2, pp. 1–17, 2018, doi: 10.3390/designs2020013.

[18] S. B. Sakri, N. B. Abdul Rashid, and Z. Muhammad Zain, “Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction”, *IEEE Access*, vol. 6, pp. 29637–29647, 2018, doi: 10.1109/ACCESS.2018.2843443.

[19] M. R. Mohebian, H. R. Marateb, M. Mansourian, M. A. Mañanas, and F. Mokarian, “A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning”, *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 75–85, 2017, doi: 10.1016/j.csbj.2016.11.004.

[20] E. Rexhepaj *et al.*, “Validation of cytoplasmic-to-nuclear ratio of survivin as an indicator of improved prognosis in breast cancer”, *BMC Cancer*, vol. 10, no. 1, p. 639, Dec. 2010, doi: 10.1186/1471-2407-10-639.

[21] A.-A. Nahid and Y. Kong, "Involvement of Machine Learning for Breast Cancer Image Classification: A Survey", *Comput. Math. Methods Med.*, vol. 2017, 2017, doi: 10.1155/2017/3781951.

[22] H. Huang *et al.*, "A new fruit fly optimization algorithm enhanced support vector machine for diagnosis of breast cancer based on high-level features", *BMC Bioinformatics*, vol. 20, 2019, doi: 10.1186/s12859-019-2771-z.

[23] P. Kaur, G. Singh, and P. Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification", *Inform. Med. Unlocked*, vol. 16, 2019, doi: 10.1016/j.imu.2019.01.001.

[24] Y. Min, X. Wei, H. Chen, K. Xiang, G. Yin, and Y. Feng, "Identifying Clinicopathological Risk Factors of the Regional Lymph Node Metastasis in Patients with T1-2Mucinous Breast Cancer: A Population-Based Study", *J. Oncol.*, vol. 2021, 2021, doi: 10.1155/2021/3866907.

[25] X. Lan *et al.*, "Application of machine learning with multiparametric dual-energy computed tomography of the breast to differentiate between benign and malignant lesions", *Quant. Imaging Med. Surg.*, vol. 12, no. 1, pp. 810–822, 2022, doi: 10.21037/qims-21-39.

[26] R. Fusco, M. Di Marzo, C. Sansone, M. Sansone, and A. Petrillo, "Breast DCE-MRI: lesion classification using dynamic and morphological features by means of a multiple classifier system", *Eur. Radiol. Exp.*, vol. 1, no. 1, 2017, doi: 10.1186/s41747-017-0007-4.

[27] R. Fusco *et al.*, "Blood oxygenation level dependent magnetic resonance imaging (Mri), dynamic contrast enhanced mri and diffusion weighted mri for benign and malignant breast cancer discrimination: A preliminary experience", *Cancers*, vol. 13, no. 10, 2021, doi: 10.3390/cancers13102421.

[28] R. Fusco *et al.*, "Radiomic and artificial intelligence analysis with textural metrics, morphological and dynamic perfusion features extracted by dynamic contrast-enhanced magnetic resonance imaging in the classification of breast lesions", *Appl. Sci. Switz.*, vol. 11, no. 4, pp. 1–16, 2021, doi: 10.3390/app11041880.

[29] M. Sansone *et al.*, "Machine Learning Approaches with Textural Features to Calculate Breast Density on Mammography", *Curr. Oncol.*, vol. 30, no. 1, pp. 839–853, 2023, doi: 10.3390/curroncol30010064.

[30] Y. Zhang *et al.*, "Risk factors for axillary lymph node metastases in clinical stage T1-2N0M0 breast cancer patients", *Med. U. S.*, vol. 98, no. 40, 2019, doi: 10.1097/MD.0000000000017481.

[31] L. Hussain *et al.*, "Machine learning classification of texture features of MRI breast tumor and peri-tumor of combined pre- and early treatment predicts pathologic complete response", *Biomed. Eng. Online*, vol. 20, no. 1, 2021, doi: 10.1186/s12938-021-00899-z.

[32] K. Wang *et al.*, "Integrated multi-omics profiling of high-grade estrogen receptor-positive, HER2-negative breast cancer", *Mol. Oncol.*, vol. 16, no. 12, pp. 2413–2431, 2022, doi: 10.1002/1878-0261.13043.

[33] Y.-M. Lei *et al.*, "Artificial Intelligence in Medical Imaging of the Breast", *Front. Oncol.*, vol. 11, 2021, doi: 10.3389/fonc.2021.600557.

[34] J. Li *et al.*, "Predicting breast cancer 5-year survival using machine learning: A systematic review", *PLoS ONE*, vol. 16, no. 4 April, 2021, doi: 10.1371/journal.pone.0250370.

[35] H. Liang, J. Li, H. Wu, L. Li, X. Zhou, and X. Jiang, "Mammographic Classification of Breast Cancer Microcalcifications through Extreme Gradient Boosting", *Electron. Switz.*, vol. 11, no. 15, 2022, doi: 10.3390/electronics11152435.

[36] G. XI *et al.*, "Nomogram model combining macro and micro tumor-associated collagen signatures obtained from multiphoton images to predict the histologic grade in breast cancer", *Biomed. Opt. Express*, vol. 12, no. 10, pp. 6558–6570, 2021, doi: 10.1364/BOE.433281.

[37] L. Zhao *et al.*, "Machine Learning Algorithms Identify Clinical Subtypes and Cancer in Anti-TIF1γ+ Myositis: A Longitudinal Study of 87 Patients", *Front. Immunol.*, vol. 13, 2022, doi: 10.3389/fimmu.2022.802499.

[38] H. Feng *et al.*, "Prediction of radiation-induced acute skin toxicity in breast cancer patients using data encapsulation screening and dose-gradient-based multi-region radiomics technique: A multicenter study", *Front. Oncol.*, vol. 12, 2022, doi: 10.3389/fonc.2022.1017435.

[39] J. Xu, X. Rao, W. Lu, X. Xie, X. Wang, and X. Li, “Noninvasive Predictor for Premalignant and Cancerous Lesions in Endometrial Polyps Diagnosed by Ultrasound”, *Front. Oncol.*, vol. 11, 2022, doi: 10.3389/fonc.2021.812033.

[40] F. Xiong, X. Cao, X. Shi, Z. Long, Y. Liu, and M. Lei, “A machine learning–Based model to predict early death among bone metastatic breast cancer patients: A large cohort of 16,189 patients”, *Front. Cell Dev. Biol.*, vol. 10, 2022, doi: 10.3389/fcell.2022.1059597.

[41] F. Su *et al.*, “Integrated Tissue and Blood miRNA Expression Profiles Identify Novel Biomarkers for Accurate Non-Invasive Diagnosis of Breast Cancer: Preliminary Results and Future Clinical Implications”, *Genes*, vol. 13, no. 11, 2022, doi: 10.3390/genes13111931.

[42] D. Gu, W. Zhao, Y. Xie, X. Wang, K. Su, and O. V. Zolotarev, “A personalized medical decision support system based on explainable machine learning algorithms and ecc features: Data from the real world”, *Diagnostics*, vol. 11, no. 9, 2021, doi: 10.3390/diagnostics11091677.

[43] A. C. Kaushik, A. Mehmood, X. Wang, D.-Q. Wei, and X. Dai, “Globally ncRNAs Expression Profiling of TNBC and Screening of Functional lncRNA”, *Front. Bioeng. Biotechnol.*, vol. 8, 2021, doi: 10.3389/fbioe.2020.523127.

[44] L. Sun *et al.*, “An image segmentation framework for extracting tumors from breast magnetic resonance images”, *J. Innov. Opt. Health Sci.*, vol. 11, no. 4, 2018, doi: 10.1142/S1793545818500141.

[45] S. Smerekanych, T. S. Johnson, K. Huang, and Y. Zhang, “Pseudogene-gene functional networks are prognostic of patient survival in breast cancer”, *BMC Med. Genomics*, vol. 13, 2020, doi: 10.1186/s12920-020-0687-0.

[46] Y. Zhang, Y. Zhou, F. Mao, R. Yao, and Q. Sun, “Ki-67 index, progesterone receptor expression, histologic grade and tumor size in predicting breast cancer recurrence risk: A consecutive cohort study”, *Cancer Commun.*, vol. 40, no. 4, pp. 181–193, 2020, doi: 10.1002/cac2.12024.

[47] L. Yang *et al.*, “PET/CT-based radiomics analysis may help to predict neoadjuvant chemotherapy outcomes in breast cancer”, *Front. Oncol.*, vol. 12, 2022, doi: 10.3389/fonc.2022.849626.

[48] Y. Peng, W. Li, and Y. Liu, “A Hybrid Approach for Biomarker Discovery from Microarray Gene Expression Data for Cancer Classification”, *Cancer Inform.*, vol. 2, p. 117693510600200, Jan. 2006, doi: 10.1177/117693510600200024.

[49] D. Ferraro *et al.*, “Microfluidic platform combining droplets and magnetic tweezers: application to HER2 expression in cancer diagnosis”, *Sci. Rep.*, vol. 6, no. 1, p. 25540, May 2016, doi: 10.1038/srep25540.

[50] N. I. Hadi, Q. Jamal, A. Iqbal, F. Shaikh, S. Somroo, and S. G. Musharraf, “Serum Metabolomic Profiles for Breast Cancer Diagnosis, Grading and Staging by Gas Chromatography-Mass Spectrometry”, *Sci. Rep.*, vol. 7, no. 1, 2017, doi: 10.1038/s41598-017-01924-9.

[51] D. Sánchez-Calderón, A. Pedraza, C. M. Urrego, A. Mejía-Mejía, A. L. Montealegre-Páez, and S. Perdomo, “Analysis of the cost-effectiveness of liquid biopsy to determine treatment change in patients with her2-positive advanced breast cancer in Colombia”, *Clin. Outcomes Res.*, vol. 12, pp. 115–122, 2020, doi: 10.2147/CEOR.S220726.

[52] S. Abbas *et al.*, “BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm”, *PeerJ Comput. Sci.*, vol. 7, p. e390, Mar. 2021, doi: 10.7717/peerj-cs.390.

[53] M. Saarela and S. Jauhiainen, “Comparison of feature importance measures as explanations for classification models”, *SN Appl. Sci.*, vol. 3, no. 2, 2021, doi: 10.1007/s42452-021-04148-9.

[54] A. H. Osman and H. M. A. Aljahdali, “An Effective of Ensemble Boosting Learning Method for Breast Cancer Virtual Screening Using Neural Network Model”, *IEEE Access*, vol. 8, pp. 39165–39174, 2020, doi: 10.1109/ACCESS.2020.2976149.

[55] M. Kaya Keleş, “Breast cancer prediction and detection using data mining classification algorithms: A comparative study”, *Teh. Vjesn.*, vol. 26, no. 1, pp. 149–155, 2019, doi: 10.17559/TV-20180417102943.

[56] C. L. Chowdhary, M. Mittal, P. Kumaresan, P. A. Pattanaik, and Z. Marszalek, “INVOLVEMENT OF MACHINE LEARNING TOOLS IN HEALTHCARE DECISION MAKING”, *Sens. Switz.*, vol. 20, no. 14, pp. 1–20, 2020, doi: 10.3390/s20143903.

[57] S. M. D. A. C. Jayatilake and G. U. Ganegoda, “Involvement of Machine Learning Tools in Healthcare Decision Making”, *J. Healthc. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6679512.

[58] N. Al-Azzam and I. Shatnawi, “Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer”, *Ann. Med. Surg.*, vol. 62, pp. 53–64, 2021, doi: 10.1016/j.amsu.2020.12.043.

[59] M. Shanbehzadeh, H. Kazemi-Arpanahi, M. Bolbolian Ghalibaf, and A. Orooji, “Performance evaluation of machine learning for breast cancer diagnosis: A case study”, *Inform. Med. Unlocked*, vol. 31, 2022, doi: 10.1016/j.imu.2022.101009.

[60] R. Massafra *et al.*, “A Clinical Decision Support System for Predicting Invasive Breast Cancer Recurrence: Preliminary Results”, *Front. Oncol.*, vol. 11, 2021, doi: 10.3389/fonc.2021.576007.

[61] Y. K. Qawqzeh, A. Alourani, and S. Ghwanmeh, “An Improved Breast Cancer Classification Method Using an Enhanced AdaBoost Classifier”, *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 1, pp. 473–478, 2023, doi: 10.14569/IJACSA.2023.0140151.

[62] M. N. Nik Ab Kadir *et al.*, “Development of Predictive Models for Survival among Women with Breast Cancer in Malaysia”, *Int. J. Environ. Res. Public. Health*, vol. 19, no. 22, 2022, doi: 10.3390/ijerph192215335.

[63] I. Ozcan, H. Aydin, and A. Cetinkaya, “Comparison of Classification Success Rates of Different Machine Learning Algorithms in the Diagnosis of Breast Cancer”, *Asian Pac. J. Cancer Prev.*, vol. 23, no. 10, pp. 3287–3297, 2022, doi: 10.31557/APJCP.2022.23.10.3287.

[64] G. Huang, Y. Yang, Y. Lei, and J. Yang, “Differences in Subjective Well-Being between Formal and Informal Workers in Urban China”, *Int. J. Environ. Res. Public. Health*, vol. 20, no. 1, p. 149, Dec. 2022, doi: 10.3390/ijerph20010149.

[65] M. B. S. Khan, Atta-Ur-Rahman, M. S. Nawaz, R. Ahmed, M. A. Khan, and A. Mosavi, “Intelligent breast cancer diagnostic system empowered by deep extreme gradient descent optimization”, *Math. Biosci. Eng.*, vol. 19, no. 8, pp. 7978–8002, 2022, doi: 10.3934/mbe.2022373.