



ANALISIS BIBLIOMETRIK PENELITIAN POHON KEPUTUSAN UNTUK PREDIKSI KANKER PAYUDARA

Suhartono^{1*}; Totok Chamidy²; Syahiduz Zaman³

^{1, 2, 3} UIN Maulana Malik Ibrahim Malang

Disubmit : 12-07-2023
Direview: 21-07-2023
Direvisi : 20-08-2023
Diterima: 25-08-2023

*Korespondensi: suhartono@ti.uin-malang.ac.id

ABSTRACT

Introduction. The purpose of this paper is to conduct a bibliometric analysis of scientific publications that discuss the use of the Decision Tree method for breast cancer prediction.

Research Methods. A total of 322 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 of the research.

Results and Discussion. The results of the analysis of the evolution article found an exponential increase in citations and the number of authors in this study in the period 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 the use of the Decision Tree method for breast cancer prediction are included in the computer field.

Conclusion. This study suggests that research on the use of the Decision Tree method for breast cancer prediction is a research topic that needs to be continuously improved.

ABSTRAK

Pendahuluan. Tujuan makalah ini adalah untuk melakukan analisis bibliometrik mengenai publikasi ilmiah yang membahas penggunaan metode Decision Tree untuk prediksi kanker payudara.

Metode Penelitian. Sebanyak 322 dokumen dari Scopus dikumpulkan untuk dianalisis menggunakan indikator bibliometrik seperti produktivitas dan sitasi. Analisis bibliometrik menghasilkan pemetaan sains berdasarkan dengan kata kunci co-occurrence, co-authorship, dan co-citation analysis untuk mencerminkan struktur konseptual, sosial, dan intelektual penelitian.

Hasil dan Pembahasan. Hasil analisis artikel yang berpengaruh menemukan peningkatan sitasi dan jumlah penulis yang eksponensial dalam penelitian ini pada periode 2005-2023, dimana China sebagai negara yang dominan dalam melakukan penelitian. Pada analisis thematic map menghasilkan tiga topik penelitian yaitu bidang kedokteran, bidang komputer dan bidang bioinformatika. Topik penelitian dalam penggunaan metode Decision Tree untuk prediksi kanker payudara adalah termasuk pada bidang komputer.

Kesimpulan dan Saran. Penelitian ini menyarankan bahwa penelitian dalam penggunaan metode Decision Tree untuk prediksi kanker payudara adalah topik penelitian yang perlu terus ditingkatkan.

Keywords: Bibliometric; Decision Tree; Breast Cancer; Prediction

1. PENDAHULUAN

Metode *Decision Tree* adalah salah satu algoritma *Machine Learning* untuk memprediksi kanker payudara pada tahap awal diagnosa dengan tingkat akurasi yang tinggi, mencapai 99% (El-Nabawy et al., 2021). Pada penelitian sebelumnya, untuk membangun model prediksi telah menggunakan berbagai metode dari algoritma *Machine Learning* lainnya, seperti *multilayer perceptron*, *naive bayes*, dan *logistic regression* (Alshammari & Mezher, 2020; Botlagunta et al., 2023).

Penggunaan algoritma *Machine Learning* dapat meningkatkan akurasi dalam prediksi dan diagnosis kanker payudara sehingga cara ini dapat membantu dalam pencegahan dan diagnosis cedera fisik (Liang et al., 2022). Selain itu, teknologi gambar MRI multiparametrik juga telah digunakan untuk klasifikasi subjenis kanker payudara (Xie et al., 2019).

Penggunaan metode *Decision Tree* telah digunakan secara luas dalam mendiagnosis kanker payudara dengan tingkat keakuratan yang tinggi. Menurut penelitian oleh Nasser dan Behadili (2022), metode ini mampu memberikan akurasi sebesar 87,83% dengan indeks Gini dan 86,77% dengan entropi.

Selain itu, penggunaan metode *Decision Tree* dengan teknik pembelajaran terawasi mampu mengidentifikasi jenis kanker dengan akurasi yang lebih tinggi dibandingkan dengan metode Logistik Regresi dan *Neural Network*. Mudunuru dan Skrzypek (2020). Oleh karena itu, Metode *Decision Tree* memiliki potensi yang besar untuk menjadi alat yang efektif dalam memprediksi dan mendiagnosis kanker payudara. Pada penelitian terbaru menunjukkan penggunaan metode *Principal Component Analysis* (PCA) untuk *feature selection* pada dataset Wisconsin Diagnostic Breast Cancer dapat meningkatkan performa pada *Decision Tree* (Chen et al., 2019).

Selain itu, sebuah metode hybrid yang menggabungkan PCA dan Boosted C5.0 *Decision Tree* dengan penalty factor diusulkan untuk mengatasi masalah ketidakseimbangan kelas dalam diagnosis kanker payudara. Metode ini telah diuji coba pada dataset kanker payudara dari UCI *Machine Learning* Repository dan terbukti efektif dalam mengatasi masalah tersebut (Tian & Zhang, 2022). Selain itu, sebuah studi menemukan bahwa metode *Decision Tree* sangat efektif dalam mendeteksi kanker payudara pada stadium awal dengan akurasi 100% pada uji coba pertama dan 97,9% pada uji coba kedua (Dobrovska & Nosovets, 2021).

Hasil penelitian oleh Assegie et al. (2021) menunjukkan metode *Decision Tree* memiliki akurasi 88,80% dalam memprediksi kanker payudara, sedangkan penelitian yang dilakukan oleh Yang et al. (2018) menunjukkan bahwa pola pewarnaan HER2-IHC dapat digunakan untuk memprediksi dan mendiagnosa kanker payudara. Penelitian lain oleh Murtaza et al. (2022) menunjukkan bahwa model prediksi dengan menggunakan algoritma *K-Nearest Neighbour* memiliki akurasi 86,24% dalam memprediksi kanker payudara.

Semakin banyaknya publikasi karya ilmiah tentang penggunaan *Decision Tree* untuk prediksi kanker payudara maka membuat kita sulit untuk mengikuti perkembangan dari semua penelitian yang diterbitkan. Selain itu, penelitian yang terfragmentasi akan menghambat kemampuan kita untuk mengumpulkan ilmu pengetahuan dan mengumpulkan bukti melalui serangkaian artikel penelitian sebelumnya. Oleh karena itu, penelitian tentang tinjauan literatur semakin penting untuk mensintesis temuan penelitian sebelumnya, mengetahui perkembangan penelitian yang terbaru (Briner & Denyer, 2012; Rousseau, 2012).

Analisis bibliometrik merupakan metode untuk menyintesis hasil penelitian dengan menggunakan metode matematika dan statistik untuk menganalisis kumpulan literatur secara sistematis dalam bidang tertentu. Namun, analisis bibliometrik mengenai penggunaan *Decision Tree* untuk prediksi kanker payudara saat sekarang ini masih terbatas. Oleh karena itu, pada artikel ini mempunyai tujuan untuk mengumpulkan literatur yang relevan dari *database* Scopus dan melakukan analisis bibliometrik untuk menjawab tiga pertanyaan penelitian terkait dengan penggunaan metode *Decision Tree*. Pertanyaan penelitian pertama adalah tren perkembangan penerapan metode *Decision Tree* (RQ 1), pertanyaan kedua adalah literatur, jurnal, penulis, dan negara yang paling berpengaruh (RQ 2), kemudian yang terakhir adalah karakteristik struktur literatur tentang *Decision Tree* (RQ 3).

2. METODE

Analisis bibliometrik adalah suatu analisis yang menggunakan metode kuantitatif untuk menganalisis kumpulan literatur ilmiah. Metode ini melibatkan analisis jumlah publikasi, kutipan, dan kutipan bersama dari topik penelitian tertentu. Pada analisis bibliometrik akan membahas analisis jumlah publikasi, kutipan, dan kutipan bersama dari berbagai artikel dan jurnal yang membahas topik ini. Sebuah tinjauan sistematis dilakukan berdasarkan pedoman PRISMA, di mana peneliti mengikuti

pernyataan sesuai dengan PRISMA 2020 (Page et al., 2021), pada penelitian ini membangun tinjauan sistematis penelitian untuk mengevaluasi penggunaan metode *Decision Tree* for prediksi kanker.

Dengan menggunakan analisis bibliometrik dan pedoman PRISMA, studi ini bertujuan untuk menyajikan tren evolusi dalam penerapan *Decision Tree* untuk prediksi kanker payudara, menganalisis literatur, jurnal, penulis, dan negara mana yang paling berpengaruh dalam penelitian ini, serta mengidentifikasi karakteristik struktur literatur tentang penggunaan *Decision Tree* untuk prediksi kanker payudara.

Penelitian ini dilakukan dengan melakukan pencarian artikel penelitian pada *database* elektronik di Scopus.Com dengan tujuan untuk melakukan analisis bibliometrik tentang penggunaan *Decision Tree* dalam prediksi kanker payudara. Dalam penelitian ini, *database* Scopus digunakan sebagai sumber data. alasan menggunakan satu basis data adalah untuk menghindari duplikasi literatur antar basis data.

Database Scopus digunakan sebagai dasar alat pencarian dokumen penelitian dan sebagai dasar evaluasi penelitian. Jumlah makalah yang dimasukkan oleh Scopus mencerminkan tingkat penelitian ilmiah dan kemampuan penelitian. Kumpulan artikel dipilih menggunakan kueri dengan dua aspek, yaitu *Technique-Related Words (Decision Tree)* dan *Cancer-Related Words (Breast Cancer)*.

Penelitian ini melakukan analisis bibliometrik berbasis *database* scopus. Beberapa analisis yang dilakukan adalah analisis jumlah makalah, analisis jumlah kutipan, analisis tren penelitian, dan analisis jurnal, penulis, negara, dan literatur yang paling berpengaruh dalam penelitian ini, seperti pada tabel 1.

Tabel 1 Keywords for research related

<i>Technique-Related Words</i>	<i>Cancer-Related Words</i>
<i>Decision Tree</i>	<i>Breast Cancer</i>

Untuk mencapai tujuan penelitian yang telah ditentukan, pencarian artikel dilakukan di *database* Scopus dengan menggunakan query keywords "BREAST CANCER" dan "DECISION TREE". Pencarian meliputi judul, abstrak, kata kunci penulis, dan kata kunci plus dalam setiap artikel yang terdapat di dalam basis data tersebut. Pencarian dilakukan dengan menggunakan kriteria inklusi dan eksklusi tertentu, seperti hanya mencari artikel dengan bahasa Inggris dan yang terbit antara tahun 2005 hingga 2023. Dalam Pencarian ini, kita dapat menemukan artikel relevan dengan topik penelitian dan dapat digunakan untuk melakukan analisis bibliometrik.

TITLE-ABS-KEY ("BREAST CANCER" AND "DECISION TREE") AND (LIMIT-TO (OA , "publisherfullgold")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) 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))

Edit Save Set alert

Gambar 1 Tampilan query pada alat pencarian di halaman Scopus

Peneliti memasukkan query sesuai dengan topik penelitian, langkah selanjutnya adalah melakukan proses download data ke file bib. Proses ini dilakukan untuk memudahkan peneliti dalam melakukan pengolahan data dan menganalisis literatur ilmiah berbasis *database*. Format file bib yang digunakan merupakan format standar untuk mengumpulkan referensi dalam penelitian akademik. Dengan mengumpulkan referensi dalam format ini, peneliti dapat melakukan analisis bibliometrik dengan mudah dan akurat, proses download di *database* scopus dapat dilihat pada gambar 2.

Select your method of export

☐ Mendeley
 ☐ ExLibris RefWorks
 ☐ SciVal
 ☐ RIS Format EndNote, Reference Manager
 ☐ CSV Excel
 ☒ BibTeX
 ☐ Plain Text ASCII In HTML

What information do you want to export?

☒ Citation information	☒ Bibliographical information	☒ Abstract & keywords	☐ Funding details	☐ Other 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	☒ Affiliations ☒ Serial identifiers (e.g. ISSN) ☒ PubMed ID ☒ Publisher ☒ Editor(s) ☒ Language of original document ☒ Correspondence address ☒ Abbreviated source title	☒ Abstract ☒ Author keywords ☒ Index keywords	☐ Number ☐ Acronym ☐ Sponsor ☐ Funding text	☐ Tradenames & manufacturers ☐ Accession numbers & chemicals ☐ Conference information ☒ Include references

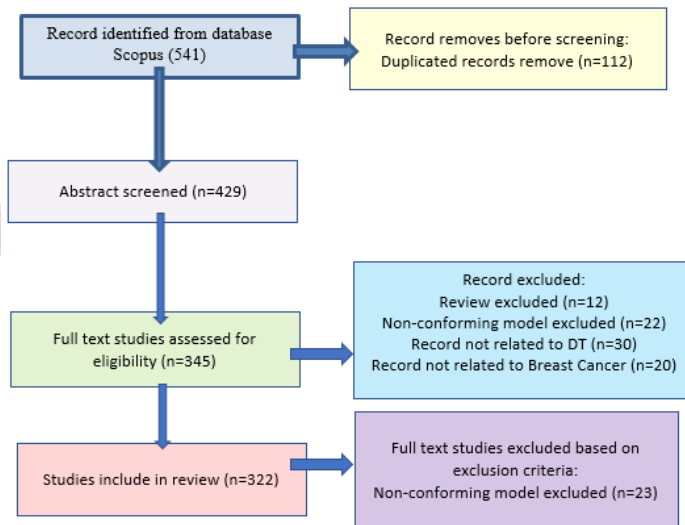
Cancel Export

Gambar 2 Tampilan download data ke file bib melalui halaman Scopus

Setelah mendapatkan file bib dari *database* scopus, maka peneliti memeriksa dan memilih artikel sesuai dengan topik penelitian menggunakan diagram alur PRISMA seperti Gambar 3.

Tahap awal adalah pencarian literatur pada *Database* scopus, di mana *query* keywords "BREAST CANCER" AND "*DECISION TREE*" digunakan untuk mencari artikel yang relevan.

Setelah itu, dilakukan seleksi awal dengan mengevaluasi judul dan abstrak, di mana artikel-artikel yang tidak relevan akan dieliminasi. Kemudian, dilakukan seleksi lanjutan dengan membaca secara penuh artikel yang masih terpilih, dan hanya artikel yang memenuhi kriteria inklusi dan eksklusi yang dipilih. Proses PRISMA ini memastikan bahwa hanya artikel-artikel yang relevan dan berkualitas tinggi yang digunakan dalam penelitian ini untuk analisis bibliometrik.



Gambar 3 Bagan alur untuk ekstraksi artikel dengan PRISMA 2020

Pada gambar 3, proses dengan alur PRISMA menghasilkan 322 artikel yang dipilih dan dimasukkan dalam analisis bibliometrik. Artikel tersebut mencakup periode dari tahun 2005 hingga Maret 2023 dan difokuskan pada penggunaan metode *Decision Tree* dalam prediksi kanker payudara. Beberapa artikel yang tidak relevan dengan topik, seperti *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) dikeluarkan dari analisis. Selanjutnya, artikel lengkap yang memenuhi kriteria inklusi dinilai dan dianalisis lebih lanjut dengan menggunakan paket R bibliometrix. Paket R bibliometrix adalah paket bahasa R yang digunakan sebagai alat untuk penelitian kuantitatif dalam bibliometrik. Paket ini memungkinkan peneliti untuk menganalisis file bib secara sistematis dan menyediakan lingkungan dan ekosistem sumber terbuka untuk analisis bibliometrik. Analisis tren, analisis statistik deskriptif, dan analisis kinerja penulis menggunakan paket R bibliometrix (<http://www.bibliometrix.org>) pada file bib dari database Scopus.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Deskripsi dari *Dataframe*

Dari proses PRISMA, peneliti mendapatkan 322 artikel terpilih, koleksi artikel tersebut dilakukan analisis deskriptif, hasil analisis deskriptif adalah informasi utama yang dikumpulkan dari database Scopus untuk 322 publikasi yang diterbitkan antara 2005 dan 2023 seperti pada tabel 2. Koleksi artikel terpilih muncul di 140 sumber yang sebagian besar adalah publikasi ilmiah jurnal sebanyak 301 artikel. Hasil analisis deskriptif menyatakan bahwa penelitian penggunaan metode *Decision Tree* untuk prediksi kanker payudara terdapat tren kenaikan sebesar 12.25%. “*Keywords Plus*” adalah total jumlah kata kunci yang sering muncul di judul artikel, hasil analisis deskriptif menyatakan bahwa jumlah “*Keywords Plus*” sebesar 2978, hasil analisa dapat dikatakan sembilan kali jumlah artikel. Masa penelitian mulai 2005 sampai 2023 (17 tahun), dan jumlah publikasi meningkat sebesar 12,25% setiap tahun seperti pada tabel 2.

Tabel 2 Hasil Analisis bibliometrik deskriptif

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2005:2023
Sources (Journals, Books, etc)	140
Documents	322
Annual Growth Rate %	12.25
Document Average Age	3.25
Average citations per doc	12.72
Average citations per year per doc	2.728
References	15284
DOCUMENT TYPES	
article	301
editorial	1
erratum	1
retracted	1
review	18
DOCUMENT CONTENTS	
Keywords Plus (ID)	2978
Author's Keywords (DE)	734
AUTHORS	
Authors	1614
Author Appearances	1785
Authors of single-authored docs	8
AUTHORS COLLABORATION	
Single-authored docs	8
Documents per Author	0.2
Co-Authors per Doc	5.54
International co-authorships %	27.95

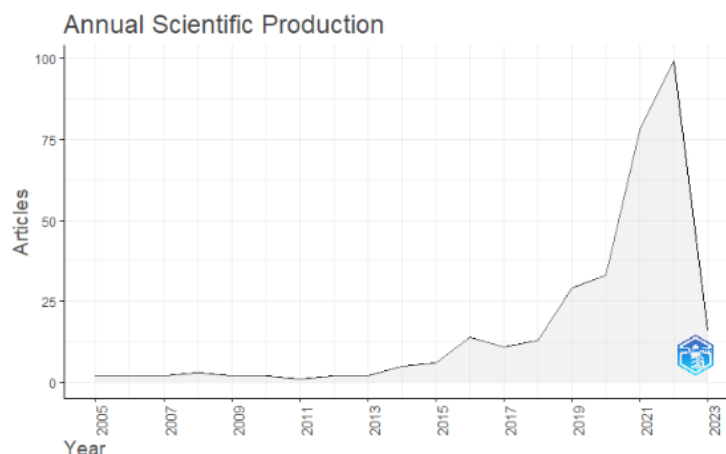
28 rows

3.2. Analisa Kinerja

Untuk menjawab RQ1 dan RQ2, kami melakukan analisis kinerja artikel dengan menggunakan beberapa indikator bibliometrik seperti jumlah publikasi, jumlah kutipan, *h-index*, *m-quotient*, dan beberapa metrik variabel berdasarkan jumlah publikasi dan kutipan. Hasil analisis menyimpulkan bahwa publikasi tentang penggunaan metode *Decision Tree* untuk prediksi kanker payudara telah meningkat secara signifikan dalam 17 tahun terakhir, dengan jumlah publikasi mencapai 322. Hasil analisis juga menghasilkan output tentang penulis, sumber, dan negara yang paling berpengaruh dalam penelitian ini, seperti penulis yang sering dikutip, jurnal yang sering diterbitkan, dan negara yang aktif dalam penelitian ini.

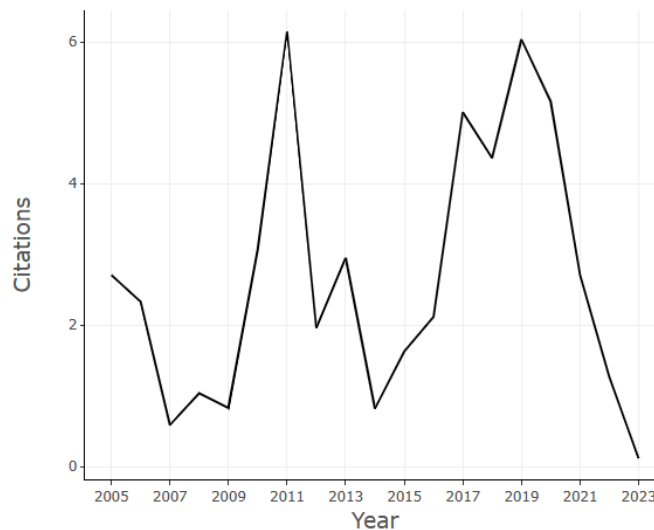
3.2.1. Analisis artikel

Sejumlah 322 artikel yang relevan dengan topik penelitian telah ditemukan dalam periode tahun 2005 hingga 2023. Dalam Gambar 4, kita dapat lihat bahwa jumlah publikasi terdapat peningkatan yang signifikan setiap tahun. Puncak tertinggi pada tahun 2022, jumlah publikasi sebesar 99 artikel. Jumlah Publikasi pernah stagnan sebesar 2 pada tahun 2005 sampai 2013. Terjadi peningkatan publikasi mulai tahun 2014 sampai tahun 2022. Peningkatan ini menunjukkan bahwa penggunaan metode *Decision Tree* semakin diminati dalam konteks prediksi kanker payudara. Semakin banyak peneliti yang tertarik untuk mengembangkan dan memperbaiki teknik ini. Hal ini menjadi dasar bagi peneliti dan praktisi dalam mengembangkan strategi baru dan terkini dalam bidang prediksi kanker payudara menggunakan metode *Decision Tree*.



Gambar 4 Distribusi jumlah publikasi setiap tahun (2005-2023)

Sedangkan untuk sitasi pada penelitian tentang penggunaan metode *Decision Tree* untuk prediksi kanker payudara menunjukkan pertumbuhan stagnan mulai tahun 2005 sampai 2023 seperti pada gambar 5, peningkatan sitasi terjadi dua kali yaitu pada tahun 2011 dan pada tahun 2019. Meski begitu, jumlah sitasi setiap tahunnya tidak menunjukkan tren tertentu, maksimal sitasi terjadi pada tahun 2011 dan 2019, jumlah maksimal sitasi adalah 6, sedangkan minimal sitasi pada tahun 2007 dan 2014, jumlah minimal sitasi adalah 1, hal ini ditunjukkan pada Gambar 5.



Gambar 5 Distribusi sitasi setiap tahun dari tahun 2005 sampai 2023

3.2.2. Analisis Jurnal Penting

Dalam analisis bibliometrik ini terdapat 322 artikel yang diterbitkan dalam 140 jurnal. Pada tabel 3 menunjukkan 10 jurnal teratas dalam hal jumlah publikasi. Pemeringkatan jurnal dalam tabel dilakukan berdasarkan jumlah publikasi (TP). Jurnal yang menempati urutan teratas adalah "Frontiers In Oncology" dengan jumlah publikasi sebanyak 19 artikel, diikuti oleh "BMC Bioinformatics" dengan 13 artikel, dan "IEEE Access" dengan 12 artikel. Pada posisi keempat dan kelima adalah "Plos One" dan "International Journal of Advanced Computer Science and Applications" dengan jumlah publikasi masing-masing sebanyak 12 artikel.

Jurnal-jurnal lain yang masuk dalam daftar 10 besar adalah "Informatics in Medicine Unlocked", "Computational and Mathematical Methods in Medicine", "Diagnostics", "BMC Cancer", dan "Scientific Report". Peringkat jurnal ini dapat menjadi acuan dalam menentukan publikasi di jurnal terkait untuk penelitian selanjutnya dalam bidang penggunaan *Decision Tree*.

Tabel 3. jurnal teratas dalam hal jumlah publikasi

	Sources <SS. Asia>	Articles <SS. Asia>
1	FRONTIERS IN ONCOLOGY	19
2	BMC BIOINFORMATICS	13
3	IEEE ACCESS	12
4	PLOS ONE	12
5	INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS	11
6	INFORMATICS IN MEDICINE UNLOCKED	10
7	COMPUTATIONAL AND MATHEMATICAL METHODS IN MEDICINE	9
8	DIAGNOSTICS	9
9	BMC CANCER	8
10	SCIENTIFIC REPORTS	8

1-10 of 10 rows

3.2.3. Analisis Artikel Penting

Analisis jumlah kutipan yang diterima oleh suatu artikel merupakan salah satu cara untuk menemukan artikel yang berpengaruh di bidangnya. Dalam kajian prediksi kanker payudara menggunakan metode *Decision Tree*, terdapat banyak artikel yang telah diterbitkan selama tahun 2005-2023. Dalam tabel 4, terdapat 10 artikel yang paling sering dikutip dalam penelitian ini dan diurutkan berdasarkan jumlah kutipan yang diterima.

Berdasarkan Tabel 4, judul artikel yang paling banyak dikutip adalah “*Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms*” dengan author Pelon et al. (2020), jumlah kutipan adalah 163 kutipan. Artikel ini memberikan kontribusi besar dalam memperdalam pemahaman tentang penggunaan *Decision Tree* dalam prediksi kanker payudara dan juga memberikan wawasan tentang peran CAF dalam metastasis kanker payudara.

Selain itu, artikel yang masuk dalam 10 artikel yang paling sering dikutip selain membahas metode *Decision Tree* juga membahas tentang berbagai teknik seperti ensemble learning, Naïve Bayes, dan Principal Component Analysis. Judul artikel dengan jumlah kutipan tertinggi kedua adalah “*Ensemble of Decision Tree reveals potential miRNA-disease associations*” dengan author Chen et al., 2019, jumlah kutipan adalah 131 kutipan. Artikel tersebut membahas tentang metode baru yang disebut *Ensemble of Decision Tree-based miRNA-disease Association Prediction* (EDTMDA) untuk memprediksi potensi asosiasi penyakit miRNA (Chen et al., 2019).

Tabel 4 Artikel yang paling banyak dikutip

Penulis (tahun)	Judul	Citations
(Pelon et al., 2020)	<i>Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms</i>	163
(Chen et al., 2019)	<i>Ensemble of Decision Tree reveals potential miRNA-disease associations</i>	131
(Ganggayah et al., 2019)	<i>Predicting factors for survival of breast cancer patients using Machine Learning techniques</i>	110
(Yue et al., 2018)	<i>Machine Learning with applications in breast cancer diagnosis and prognosis</i>	107
(Sakri et al., 2018)	<i>Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction</i>	104
(Mohebian et al., 2017)	<i>A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning</i>	99
(Rexhepaj et al., 2010)	<i>Validation of cytoplasmic-to-nuclear ratio of survivin as an indicator of improved prognosis in breast cancer</i>	80
(Nahid & Kong, 2017)	<i>Involvement of Machine Learning for Breast Cancer Image Classification: A Survey</i>	73
(Huang et al., 2019)	<i>A new fruit fly optimization algorithm enhanced support vector machine for diagnosis of breast cancer based on high-level features</i>	65
(Kaur et al., 2019)	<i>Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification</i>	61

3.2.4. Analisis Penulis Penting

Data dalam bentuk teks yang tidak terstruktur seperti artikel atau dokumen memiliki struktur yang terbatas atau tidak ada, bentuk ini mengakibatkan komputer mengalami kesulitan untuk mengolah semantiknya. Namun, dengan adanya analisis bibliometric maka data teks tersebut dapat dianalisis dengan cara mengekstrak metadata yang mewakili fitur-fiturnya. Fitur-fitur tersebut dapat diubah menjadi bentuk terstruktur yang merupakan representasi perantara dari dokumen.

Berdasarkan Tabel 5, author Kaur et al., 2019 merupakan penulis kolaboratif tertinggi pertama. Author Kaur memproduksi satu artikel relevan dengan topik, dan artikel nya menerima 132 kutipan. Dalam studi ini, Kaur et al., 2019 mengusulkan pendekatan baru untuk diagnosis dan prediksi kanker payudara menggunakan metode *Deep Learning* (DL) dan *Support Vector Machine* (SVM) dengan

akurasi rata-rata 95%, 94%, dan 98% untuk tiga kelas, serta sensitivitas dan spesifisitas yang lebih tinggi dibandingkan metode lainnya. (Kaur et al., 2019).

Chen H adalah peneliti kolaboratif dengan Yue et al., 2018, Chen H penulis kolaboratif dengan jumlah artikel yang dikutip tertinggi kedua, author Chen H menghasilkan 4 artikel relevan. Dia mengusulkan “*Machine Learning with applications in breast cancer diagnosis and prognosis*” dan menerima lebih dari 107 kutipan. Penelitian membahas tentang kanker payudara dan diagnosis dini serta klasifikasi akurat pasien ke dalam kelompok ganas atau jinak sangat penting (Yue et al., 2018).

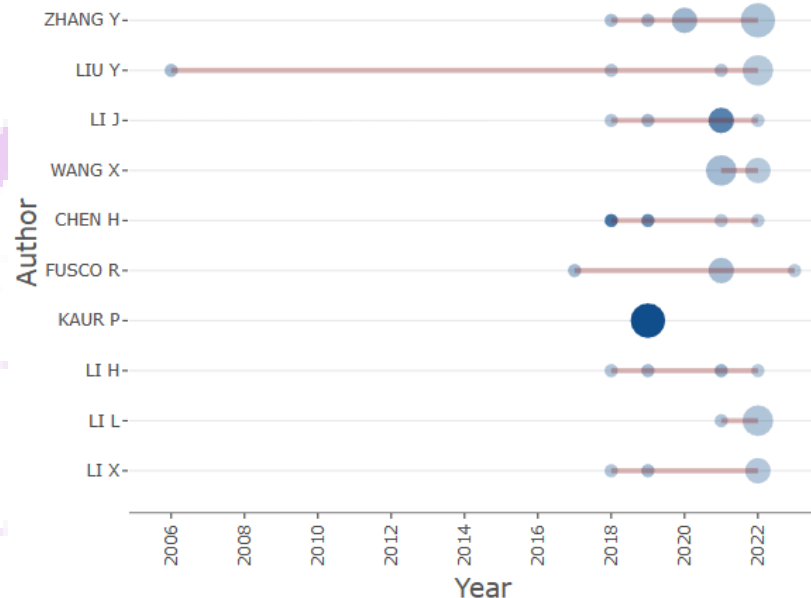
Tabel 5 Peringkat penulis paling produktif

Author	year	freq	TC	TCpY	Author
Chen H	2018	1	107	17,833	(Yue et al., 2018)
	2019	1	65	13	(Huang et al., 2019)
	2021	1	0	0	(Min et al., 2021)
	2022	1	0	0	(Lan et al., 2022)
Fusco R	2017	1	24	3,429	(Fusco et al., 2017)
	2021	2	8	2,667	(Fusco et al., 2021) ;(Fusco et al., 2021)
	2023	1	0	0	(Sansone et al., 2023)
Kaur P	2019	1	132	26,4	(Kaur et al., 2019)
Li H	2018	1	6	1	(Yang et al., 2018)
	2019	1	16	3,2	(Zhang et al., 2019)
	2021	1	16	5,333	(Hussain et al., 2021)
	2022	1	0	0	(Wang et al., 2022)
Li J	2018	1	6	1	(Yang et al., 2018)
	2019	1	16	3,2	(Zhang et al., 2019)
	2021	2	46	15,333	(Li et al., 2021);(Lei et al., 2021)
	2022	1	0	0	(Liang et al., 2022)
Li L	2021	1	3	1	(XI et al., 2021);(Wang et al., 2022);(Zhao et al., 2022)
	2022	3	2	1	(Wang et al., 2022)
Li X	2018	1	6	1	(Yang et al., 2018)
	2019	1	16	3,2	(Zhang et al., 2019)
	2022	2	1	0,5	(Feng et al., 2022);(Xu et al., 2022)
Liu Y	2006	1	53	2,944	(Alromema et al., 2023)
	2018	1	6	1	(Yang et al., 2018)
	2021	1	3	1	(XI et al., 2021)
	2022	2	0	0	(Xiong et al., 2022);(Su et al., 2022)
Wang X	2021	2	9	3	(Xu et al., 2022);(Lan et al., 2022)
	2022	3	0	0	(Gu et al., 2021);(Kaushik et al., 2021);(Jiang et al., 2021)
Zhang Y	2018	1	15	2,5	(Sun et al., 2018)
	2019	1	16	3,2	(Zhang et al., 2019);(Chen et al., 2019)
	2020	2	16	4	(Smerekanych et al., 2020);(Zhang et al., 2020)
	2022	4	3	1,5	(Li et al., 2022);(Yang et al., 2022);(Su et al., 2022)

Dalam analisis bibliometrik pada penelitian ini masih ditemukan artikel dengan jumlah kutipan rendah (0 kutipan), menurut peneliti alasan utama rendahnya jumlah kutipan adalah kurangnya pemahaman tentang penelitian kolaborasi di dalam dan di luar negeri, kurangnya inovasi, atau mengabaikan referensi sejawat dalam negeri.

Penulis juga membagi satu artikel menjadi dua atau lebih untuk menambah jumlah publikasi pada *author*, yang dapat mempengaruhi kualitas makalah. Untuk menganalisis kinerja penulis dan distribusi tahunan jumlah publikasi diunggah dalam Gambar 6.

Dari 10 penulis dengan jumlah publikasi tertinggi menunjukkan tren penelitian yang terus naik ditunjukkan 9 penulis masih menciptakan artikel. Ukuran lingkaran pada grafik menunjukkan jumlah makalah yang diterbitkan oleh penulis, dan semakin besar lingkaran, semakin banyak makalah yang diterbitkan pada tahun tersebut. Selain itu, warna lingkaran mewakili berapa kali penulis telah dikutip, dan semakin gelap warna lingkaran, semakin tinggi jumlah kutipan yang diterima oleh penulis pada tahun tersebut. Author Kaur P menerbitkan 1 artikel pada tahun 2019 dengan jumlah kutipan 132 kutipan per tahun, jumlah penerbitan terkecil tetapi jumlah kutipan terbanyak.



Gambar 6. Produksi penulis top dari waktu ke waktu.

Author Liu Y adalah perintis pertama yang memperkenalkan penggunaan metode ini untuk prediksi kanker payudara pada tahun 2006. Liu Y menerbitkan sebuah artikel di jurnal *Cancer INFORMATICS*, yang membahas pendekatan hibrida untuk menemukan biomarker kanker dari data ekspresi gen (Peng et al., 2006). Liu Y juga membandingkan enam metode filter dan tiga metode pembungkus untuk menentukan biomarker dan mengembangkan pendekatan hibrida yang menggabungkan kedua metode tersebut. Author Fusco R menjadi pionir kedua yang menerbitkan artikel tentang prediksi kanker payudara menggunakan fitur dinamis dan morfologis dengan beberapa pengklasifikasi (Fusco et al., 2017).

3.3. Pemetaan Sains

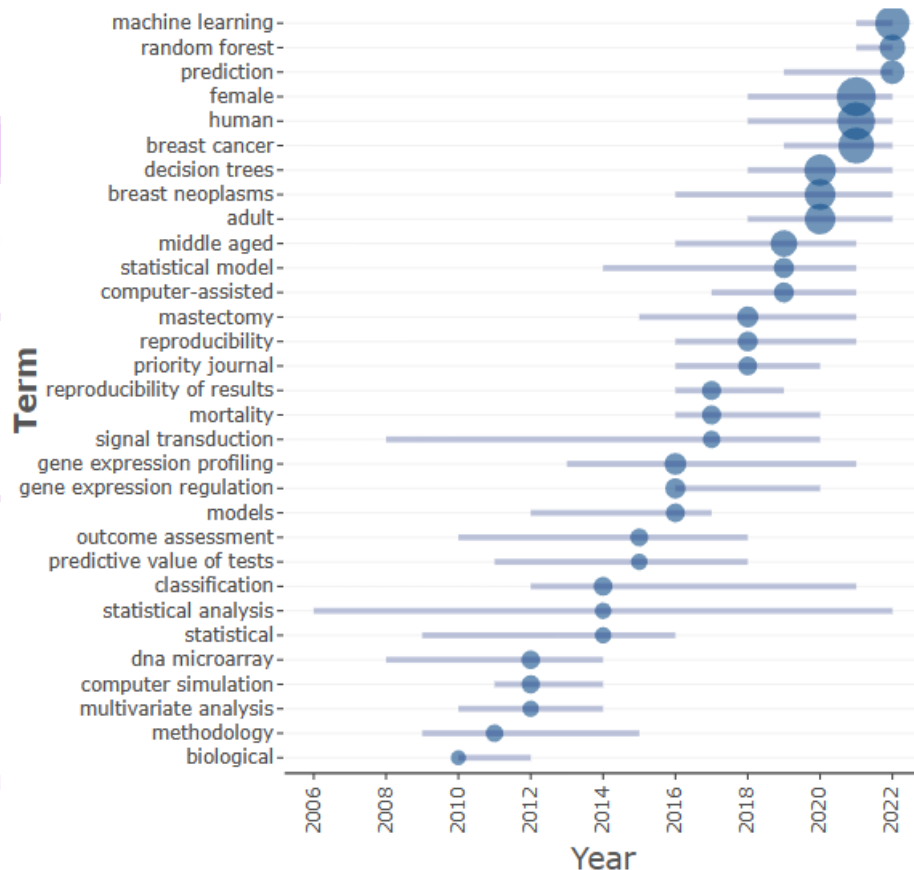
Analisis pemetaan ilmiah menjelaskan bagaimana struktur konseptual, *co-occurrence network*, dan intelektual literatur tersebut dicirikan. Pemetaan ilmiah akan mengungkapkan informasi tentang kelompok topik yang dibahas, tren artikel yang sering disitasi, keywords yang sering digunakan. Selain itu, analisis ini akan mengevaluasi kolaborasi antara penulis, serta mengidentifikasi penulis dan jurnal yang paling berpengaruh.

3.3.1. Analisis Struktur Konsep

Artikel adalah salah satu laporan ilmiah, artikel memainkan peran penting dalam penyampaian informasi ilmiah. KeyWords dalam artikel dapat membantu dalam pencarian dan penyimpanan suatu artikel. Selain itu, KeyWords Plus dapat digunakan untuk menambah pencarian artikel berdasarkan

kata kunci atau judul. Untuk membentuk jaringan co-occurrence KeyWords Plus dapat menggunakan fungsi 'biblioNetwork' dalam paket 'bibliometrix'.

Dalam gambar 7, Jumlah dan *ranking* kata kunci yang sering muncul dalam kumpulan artikel dapat memberikan gambaran tentang topik penelitian. Oleh karena itu, KeyWords dan KeyWords Plus sangat penting dalam analisis bibliometrik dan dapat membantu peneliti dalam menjelajahi kumpulan artikel terkait topik tertentu. Kata kunci *Breast Cancer* dan *Decision Tree* adalah saling berdekatan, sehingga dapat dikatakan bahwa dua keywords ini adalah sering dipakai bersama dalam penjelasan pada artikel, *Keyword Decision Tree* menempati urutan ke 6 yang sering dipakai pada kumpulan artikel dan mulai digunakan untuk penelitian mulai tahun 2016.



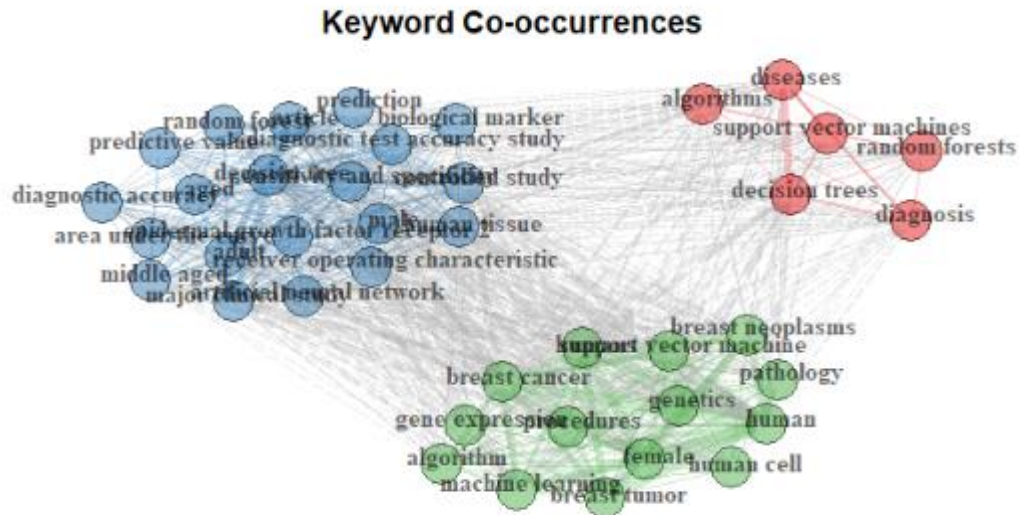
Gambar 7 Kata kunci penelitian yang sering muncul

3.3.2. Analisis Jaringan Co-Occurrence

Dalam analisis jaringan *co-occurrence* dari kata kunci plus terdapat tiga kelompok utama dalam jaringan pada kumpulan artikel ini seperti pada gambar 8. Kelompok pertama terdiri dari kata kunci yang berhubungan dengan konsep kanker payudara pada bidang kedokteran seperti "*breast cancer*", "*mammary carcinoma*", dan "*breast neoplasm*". Kelompok ini ditandai dengan lingkaran hijau.

Sedangkan kelompok kedua terdiri dari kata kunci yang berkaitan dengan metode analisis kanker payudara pada bidang komputer seperti "*Decision Tree*", "*Machine Learning*", dan "*statistical analysis*". Kelompok ini ditandai dengan lingkaran merah.

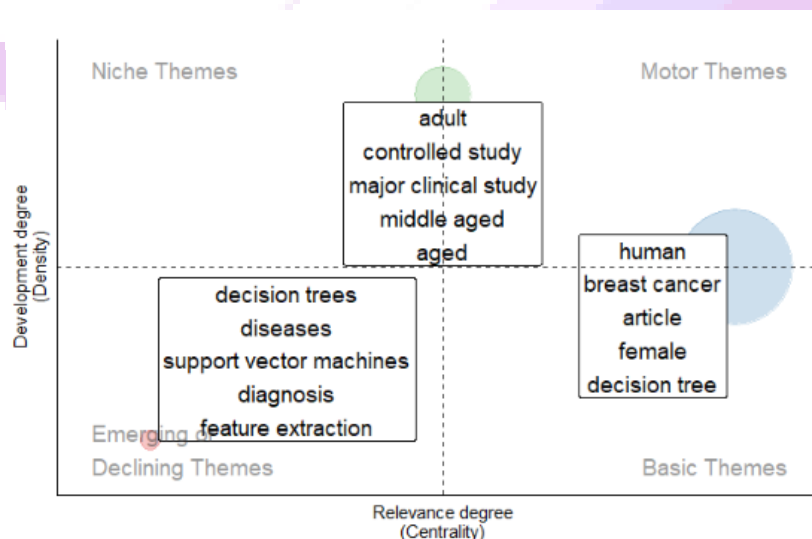
Selain itu, terdapat juga beberapa kata kunci yang muncul secara terpisah dari kedua kelompok utama seperti "*mammary carcinoma*", dan "*breast neoplasm*", "*predictive model*", "*diagnostic*" dan "*genetic algorithm*". Kelompok ini yang berhubungan dengan konsep kanker payudara pada bidang bioinformatika. Kelompok ini ditandai dengan lingkaran biru.



Gambar 8 Jaringan kejadian bersama kata kunci plus

3.3.3. Analisis Struktur Pemetaan

Dalam rangka mempermudah identifikasi artikel yang terkait dengan penelitian penggunaan *Decision Tree* pada prediksi kanker payudara maka analisis struktur pemetaan dapat digunakan. Pada conceptual structure map ini terdapat 250 kata kunci yang muncul dalam kumpulan artikel. Dari 734 keywords dan 2974 keywords plus diambil 250 kata kunci. 250 kata kunci dikategorikan menjadi tiga kelompok tema penelitian. Kelompok pertama tema penelitian kanker payudara pada bidang kedokteran. Kelompok ini ditandai dengan lingkaran hijau. Sedangkan kelompok kedua tema penelitian kanker payudara pada bidang kedokteran. Kelompok ini ditandai dengan lingkaran merah. Kelompok ketiga tema penelitian kanker payudara pada bidang bioinformatika. Kelompok ini ditandai dengan lingkaran biru. Hasil pengelompokan tersebut kemudian disajikan dalam bentuk diagram strategis pada Gambar 9. Diagram tersebut dapat digunakan untuk memvisualisasikan hubungan antara kata kunci dan tema penelitian, sehingga mempermudah pengguna dalam menemukan artikel terkait dengan topik penelitian.



Gambar 9. Peta tematik berdasarkan kata kunci

Pada Gambar 9 menjelaskan diagram strategis yang terdiri dari tiga topik penelitian berdasarkan kumpulan artikel. Setiap topik penelitian dalam diagram strategis memiliki dua parameter, yaitu *Centrality* dan *Density*. *Centrality* diukur oleh sumbu horizontal dan merupakan kekuatan koneksi eksternal antara topik penelitian dengan topik penelitian pada kumpulan artikel. Parameter ini digunakan untuk mengukur pentingnya topik penelitian dalam pengembangan penelitian. Sementara itu, *Density* diukur oleh sumbu vertikal dan merupakan kekuatan koneksi antara kata kunci dalam topik penelitian. *Density* dapat digunakan untuk mengukur derajat topik penelitian.

Melalui peta ini, dapat diketahui bahwa tema penelitian kanker payudara pada bidang komputer memiliki *Centrality* dan *Density* yang rendah, sementara tema penelitian kanker payudara pada bidang kedokteran memiliki *Density* yang tinggi, dan kanker payudara pada bidang bioinformatika memiliki *Centrality* yang tinggi. Diagram strategis pada Gambar 9 juga dapat memberikan informasi tentang posisi masing-masing tema penelitian dalam pengembangan bidang penelitian kanker payudara. Kuadran kanan atas dari diagram strategis menunjukkan bahwa tema-tema yang berhubungan dengan penelitian kanker payudara memiliki kepadatan dan sentralitas yang tinggi, sehingga tema-tema tersebut dikembangkan dengan baik dan penting untuk konstruksi bidang penelitian.

Kuadran kanan bawah menampilkan tema penelitian dengan sentralitas yang tinggi tetapi kepadatan rendah, menunjukkan bahwa tema penelitian tersebut penting untuk pengembangan bidang penelitian, tetapi tidak berkembang dengan baik, dan umumnya merupakan tema dasar dalam bidang penelitian. Tema dengan kerapatan tinggi tetapi sentralitas rendah ditemukan di kuadran kiri atas, menunjukkan bahwa tema penelitian tersebut telah berkembang dengan baik tetapi memiliki dampak yang terbatas pada bidang penelitian.

Sementara kuadran kiri bawah menyoroti tema dengan kerapatan dan sentralitas yang rendah, yang dapat dianggap sebagai tema yang kurang penting dalam bidang penelitian. Analisis peta tematik pada gambar 9 menunjukkan penelitian penggunaan metode *Decision Tree* untuk prediksi kanker payudara memiliki tiga topik penelitian utama. Meskipun topik penelitian kanker payudara pada bidang komputer menunjukkan densitas rendah dan centrality rendah, tetapi topik penelitian ini tetap dianggap penting untuk dikembangkan karena memiliki potensi untuk menjadi pusat pengembangan seluruh penelitian bidang kanker payudara. Yang kedua meskipun jumlah artikel penelitian yang terkait dengan bidang komputer masih sedikit dibandingkan dengan bidang kedokteran dan bidang informatika, berdasarkan analisis *thematic map*, bidang komputer termasuk adalah sedang berkembang dan memerlukan perhatian dalam pengembangannya.

3.3.4. Analisis Kontribusi Artikel

Berikut adalah 15 artikel yang paling berkontribusi pada masing-masing topik penelitian (5 artikel) seperti pada table 7, terdapat lima artikel dengan kontribusi tinggi di topik penelitian kedokteran (warna hijau) adalah (Pelon et al., 2020), (Ferraro et al., 2016), (Rexhepaj et al., 2010), (Hadi et al., 2017), and (Sánchez-Calderón et al., 2020). Artikel Pelon et al. (2020) menggunakan heterogenitas fibroblast terkait kanker di kelenjar getah bening aksila, sedangkan artikel Ferraro et al. (2016) menggunakan platform mikrofluida yang menggabungkan tetesan cairan dan pinset magnetik. Artikel Rexhepaj et al. (2010) menggunakan rasio sitoplasma ke inti dari survivin terkonfirmasi sebagai indikator prognosis, sedangkan artikel Hadi et al. (2017) menggunakan profil Metabolomik Serum untuk Diagnosis, Klasifikasi, dan Stadium Kanker Payudara. Sementara itu, artikel Sánchez-Calderón et al. (2020) menggunakan analisis biaya-manfaat dari biopsi cair untuk menentukan perubahan pengobatan pada pasien.

Tabel 7 artikel yang paling berkontribusi pada masing-masing topik

Tema	Tahun	Kontribusi	Pengarang, Tahun
Penulis menggunakan heterogenitas fibroblast terkait kanker di kelenjar getah bening aksila	2020	40,75	(Pelon et al., 2020)
Penulis menggunakan platform mikrofluida yang menggabungkan tetesan cairan dan pinset magnetik	2016	5,625	(Ferraro et al., 2016)
Penulis menggunakan rasio sitoplasma ke inti dari survivin terkonfirmasi sebagai indikator prognosis	2010	2,929	(Rexhepaj et al., 2010)
Penulis menggunakan profil Metabolomik Serum untuk Diagnosis, Klasifikasi, dan Stadium Kanker Payudara	2017	5,714	(Hadi et al., 2017)
Peneliti menggunakan analisis biaya-manfaat dari biopsi cair untuk menentukan perubahan pengobatan pada pasien	2020	3,25	(Sánchez-Calderón et al., 2020)
Peneliti menggunakan optimasi swarm partikel untuk seleksi fitur dalam prediksi kambuhnya kanker payudara.	2018	17,333	(Sakri et al., 2018)
Peneliti menggunakan pendekatan BCD-WERT untuk deteksi kanker payudara menggunakan fitur efisien.	2021	18,667	(Abbas et al., 2021)
Peneliti menggunakan perbandingan ukuran kepentingan fitur sebagai penjelas untuk model klasifikasi.	2021	18,333	(Saarela & Jauhiainen, 2021)
Peneliti menggunakan <i>Ensemble Boosting Learning Method</i> efektif digunakan untuk <i>virtual screening</i> kanker payudara menggunakan model <i>Neural Network</i> .	2020	7,25	(Osman & Aljahdali, 2020)
Peneliti menggunakan dalam studi perbandingan ini, digunakan algoritma klasifikasi data mining untuk memprediksi dan mendeteksi kanker payudara.	2019	5,4	(Kaya Keleş, 2019)
Peneliti menggunakan <i>Ensemble</i> dari <i>Decision Tree</i> untuk mengungkap potensi asosiasi miRNA-penyakit.	2019	26,2	(Chen et al., 2019)
Peneliti menggunakan sistem segmentasi dan klasifikasi efisien pada gambar medis menggunakan pengelompokan <i>Intuitionist Possibilistic Fuzzy C-mean</i> dan algoritma <i>Fuzzy SVM</i> .	2020	12,75	(Chowdhary et al., 2020)
Peneliti menggunakan penggunaan alat pembelajaran mesin dalam pengambilan keputusan kesehatan.	2021	14,667	(Jayatilake & Ganegoda, 2021)
Peneliti menggunakan perbandingan Model Pembelajaran Mesin Berawasi dan Semi-Berawasi dalam Mendiagnosis Kanker Payudara.	2021	11,333	(Al-Azzam & Shatnawi, 2021)

Lima artikel lainnya yang paling mewakili tema penelitian dengan bidang komputer (warna merah) adalah (Sakri et al., 2018), (Abbas et al., 2021), (Saarela & Jauhiainen, 2021), (Osman & Aljahdali, 2020), and (Kaya Keleş, 2019). Artikel Sakri et al. (2018) menggunakan optimasi swarm partikel untuk seleksi fitur dalam prediksi kambuhnya kanker payudara, sedangkan artikel Abbas et al. (2021) menggunakan pendekatan BCD-WERT untuk deteksi kanker payudara menggunakan fitur efisien. Artikel Saarela & Jauhiainen (2021) menggunakan perbandingan ukuran kepentingan fitur sebagai penjelas untuk model klasifikasi, sedangkan artikel Osman & Aljahdali (2020) menggunakan

Ensemble Boosting Learning Method efektif digunakan untuk *virtual screening* kanker payudara menggunakan model *Neural Network*. Sementara itu, artikel Kaya Keleş (2019) menggunakan algoritma klasifikasi data mining untuk memprediksi dan mendeteksi kanker payudara. Lima artikel lainnya yang paling mewakili bidang bioinformatika (warna biru) adalah (Chen et al., 2019), (Chowdhary et al., 2020), (Jayatilake & Ganegoda, 2021), (Jayatilake & Ganegoda, 2021), and (Al-Azzam & Shatnawi, 2021).

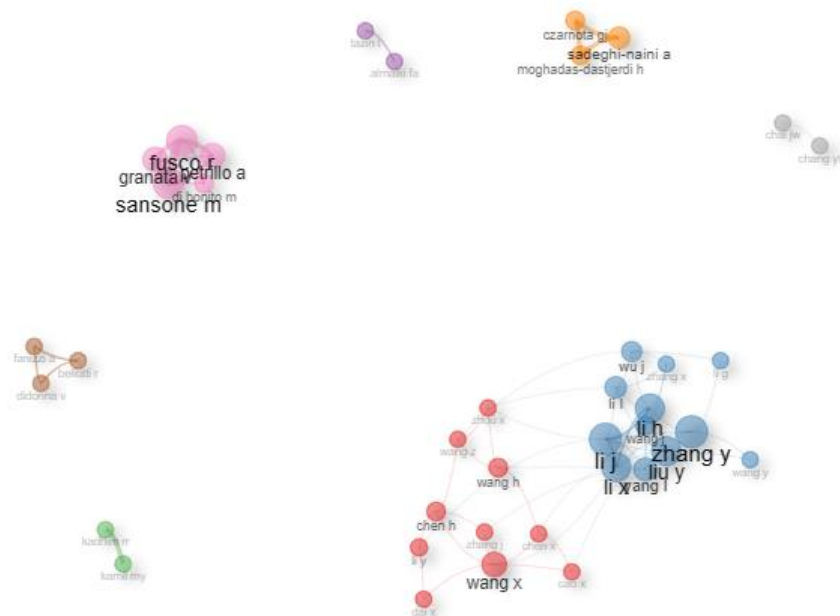
Pada topik ini, para peneliti terus mengembangkan pendekatan bidang bioinformatika untuk memperbaiki diagnosis dan pengobatan penyakit, terutama kanker payudara. Chen et al. (2019) menggunakan *Ensemble* dari *Decision Tree* untuk mengungkap potensi asosiasi miRNA-penyakit, sedangkan Chowdhary et al. (2020) menggunakan sistem segmentasi dan klasifikasi efisien pada gambar medis menggunakan pengelompokan *Intuitionist Possibilistic Fuzzy C-mean* dan algoritma *Fuzzy SVM*. Selain itu, Jayatilake & Ganegoda (2021) memanfaatkan alat pembelajaran mesin untuk membantu dalam pengambilan keputusan kesehatan, sementara Al-Azzam & Shatnawi (2021) membandingkan Model Pembelajaran Mesin Berawasi dan Semi-Berawasi dalam Mendiagnosis Kanker Payudara.

Dalam rangka memperoleh informasi lebih lanjut tentang kontribusi makalah terkait dengan tiga topik pada bidang penelitian, Tabel 7 dapat memperlihatkan lima artikel dengan kontribusi terbesar untuk masing-masing topik. Topik pertama, yang terkait dengan bidang komputer, makalah dengan kontribusi besar adalah "*A novel hybrid feature selection approach using Decision Tree and artificial bee colony algorithm for breast cancer diagnosis*" oleh Sakri et al. (2018).

Sementara itu, untuk klaster kedua, terkait dengan bidang bioinformatika, makalah dengan kontribusi besar adalah "*Artificial intelligence for breast cancer detection in mammography: A systematic review*" oleh Chowdhary et al. (2020). Makalah terkait topik ketiga, yaitu bidang kedokteran, mencakup "*Temporal patterns of radiotherapy-induced skin toxicity among breast cancer patients*" oleh Wang et al. (2017).

3.3.5. Tatanan Sosial

Pengelompokan terhadap kolaborasi penulis menggunakan Algoritma Louvain. Hasil Algoritma Louvain berbentuk visualisasi, visualisasi dapat menjelaskan hubungan antara penulis dalam kerjasama penulisan artikel seperti dalam gambar 8. Pada visualisasi terdapat 30 penulis paling berpengaruh yang bekerja sama dalam penelitian sesuai dengan topik selama 2005-2023. Visualisasi ini memberikan wawasan tentang kolaborasi penelitian di antara penulis dan membantu dalam memahami jaringan kerja ilmiah.

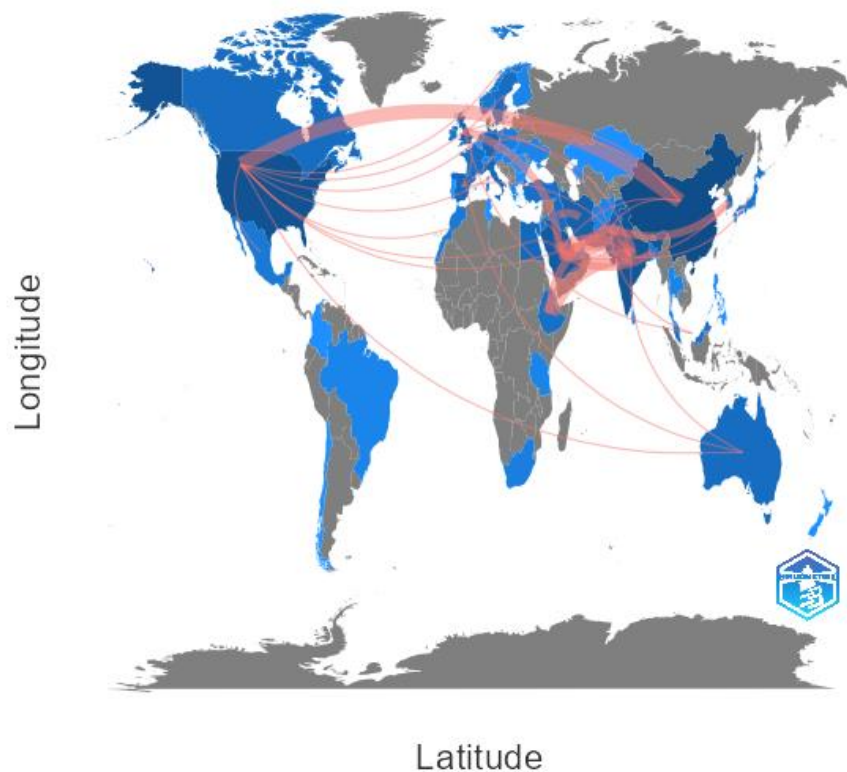


Gambar 8 Struktur sosial kumpulan data.

Pada gambar 8, ukuran lingkaran menunjukkan jumlah artikel yang ditulis oleh penulis dan cakupan silang antar lingkaran menunjukkan jumlah artikel yang ditulis bersama antara penulis. Perlu dicatat bahwa simpul terisolasi tidak berarti bahwa penulis ini tidak berkolaborasi dengan orang lain, tetapi hanya membuktikan bahwa mereka tidak berkolaborasi dengan 30 penulis paling berpengaruh teratas. Oleh karena itu, untuk menghindari salah tafsir, simpul-simpul yang terisolasi ini dihilangkan dari gambar. Seperti yang terlihat dari gambar, ada 8 grup kolaboratif dari 30 penulis berpengaruh teratas, yang terbagi dalam 3 kategori. Kelompok kolaboratif yang dipimpin oleh LIU Y memiliki hubungan kolaboratif yang paling banyak dengan ditunjukkan warna biru gelap.

Visualisasi adalah menjelaskan tentang kolaborasi penulis melalui teknik co-occurrence seperti ditunjukkan pada Gambar 8. Ukuran lingkaran menunjukkan jumlah artikel yang ditulis oleh penulis, dan cakupan silang antar lingkaran menunjukkan jumlah artikel yang ditulis bersama oleh penulis. Perlu dicatat bahwa simpul terisolasi tidak berarti bahwa penulis tidak berkolaborasi dengan orang lain, tetapi hanya membuktikan bahwa mereka tidak berkolaborasi dengan 30 penulis berpengaruh teratas. Seperti yang terlihat pada gambar 8, Visualisasi terdapat 8 grup kolaboratif di antara 30 penulis berpengaruh teratas. Kelompok kolaboratif yang dipimpin oleh Liu Y memiliki hubungan kolaboratif yang paling sering ditunjukkan dengan warna biru. Gambar 9 menunjukkan asal negara dari penulis artikel tentang penggunaan *Decision Tree* untuk prediksi kanker payudara di seluruh dunia.

Lingkaran besar merepresentasikan negara yang diwakili dengan jumlah korespondensi nasional, sedangkan ketebalan hubungan antar negara menunjukkan frekuensi kolaborasi antar negara tersebut. Dapat dilihat bahwa Amerika Serikat dan Cina menjadi negara dengan warna yang paling gelap, menandakan bahwa para peneliti dari negara-negara ini telah menulis artikel dengan jumlah terbanyak, dan sesuai dengan hasil analisis. Hubungan yang paling sering terjadi adalah antara AS dan Cina, menunjukkan bahwa peneliti dari kedua negara lebih sering berkolaborasi dan berkomunikasi. Selain itu, India dan Iran juga menempati posisi kedua dalam frekuensi kolaborasi, sebagaimana terlihat pada gambar 9.

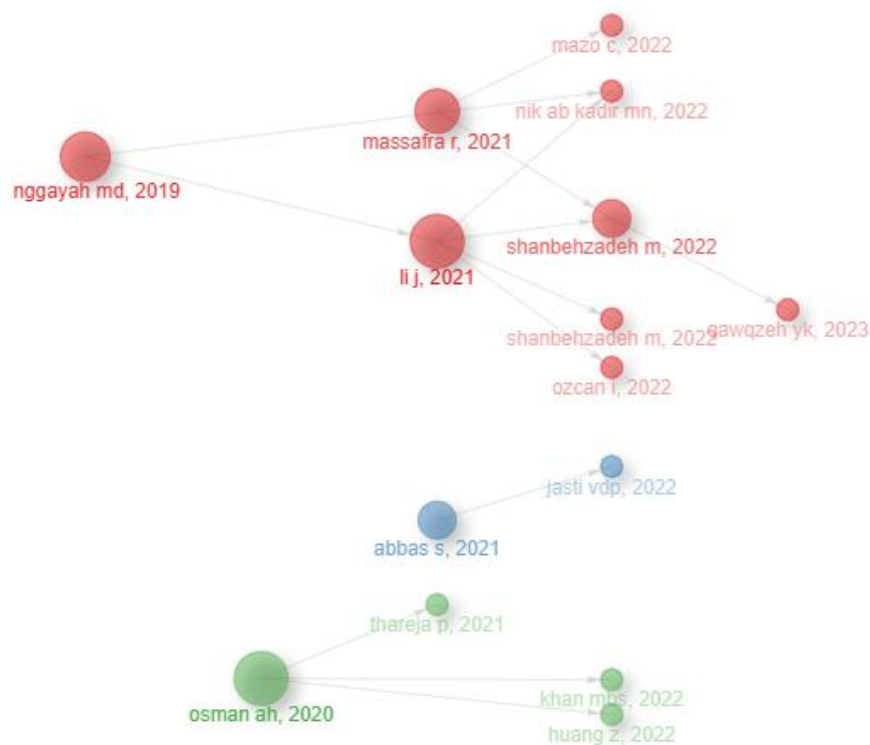


Gambar 9 Tingkat kolaborasi antar negara berdasarkan afiliasi penulis.

3.3.6. Analisis Struktur Intelektual

Gambar 10 menampilkan historiografi penelitian tentang penggunaan metode *Decision Tree* untuk prediksi kanker payudara berdasarkan jaringan kutipan. Dalam 20 artikel yang dianalisis mulai dari artikel yang ditulis oleh Ganggayah et al. pada tahun 2019 hingga artikel Qawqzeh et al. pada tahun 2023, terdapat tiga kelompok jaringan kutipan dengan warna yang berbeda, yaitu biru, merah, dan hijau.

Klaster pertama dengan font warna merah dengan 9 penulis terdiri dari penulis Shanbehzadeh et al. pada tahun 2022, Li et al. pada tahun 2021, Massafra et al. pada tahun 2021, Ganggayah et al., pada tahun 2019, Qawqzeh et al. pada tahun 2023, Nik Ab Kadir et al. pada tahun 2022, Ozcan et al. pada tahun 2022, dan Mortazavi et al. pada tahun 2022. Klaster kedua dengan font warna merah dengan 2 penulis terdiri dari Abbas et al. pada tahun 2021 dan Zhang et al. pada tahun 2019. Klaster ketiga dengan font warna hijau dengan 4 penulis, penulis Osman & Aljahdali pada tahun 2020, penulis Huang & Chen pada tahun 2022, penulis Thareja & Chhillar tahun 2021, dan Khan et al. pada tahun 2022. Dari gambar 10, dapat dilihat hubungan kutipan antara penulis-penulis yang terkait dalam penelitian ini selama kurun waktu 2019-2023.



Gambar 10 Evolusi bersejarah dari kutipan bersama di antara 20 artikel paling relevan.

Gambar 10 menunjukkan perkembangan sejarah dari 20 artikel berpengaruh tentang prediksi kanker payudara sesuai dengan urutan kronologis dan klaster. Pada klaster pertama, terdapat author paling berpengaruh pertama yaitu Ganggayah dengan judul artikel “*Predicting factors for survival of breast cancer patients using Machine Learning techniques*”. Artikel ini mempunyai *local citation score* (LCS) sebesar 3 dan *global citation score* (GCS) sebesar 110, yang diterbitkan oleh BMC Medical Informatics and Decision Making pada tahun 2019 seperti pada tabel 8.

Pada penelitian ini membangun model prediksi menggunakan *Decision Tree*, *random forest*, *neural networks*, *extreme boost*, *logistic regression*, dan *support vector machine*. Variabel penting yang diidentifikasi termasuk klasifikasi tahap kanker, ukuran tumor, jumlah total kelenjar getah bening aksila yang diangkat, jumlah kelenjar getah bening positif, jenis pengobatan primer, dan metode diagnosis. Seluruh algoritma menghasilkan akurasi yang mirip, dan metode *Machine Learning* dapat menjadi alat prediksi alternatif (Ganggayah et al., 2019).

Tabel 8 Tabel informasi 20 artikel untuk simpul-simpul penting jaringan kutipan.

Label <chr>	Year <abi>	LCS <abi>	GCS <abi>
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

Artikel yang ditulis oleh Ganggayah et al pada tahun 2019 telah menjadi dasar kutipan artikel lain pada tahun-tahun berikutnya, termasuk artikel yang ditulis oleh Li J pada tahun 2021. Artikel ini menjelaskan penelitian terdahulu yang menggunakan *Machine Learning* (ML) untuk memprediksi tingkat kelangsungan hidup pasien kanker payudara selama lima tahun, temuan yang didapat bahwa terdapat 31 artikel yang memenuhi kriteria inklusi, sebagian besar diterbitkan setelah tahun 2013 dan menggunakan metode ML yang paling sering digunakan seperti *Decision Trees*, *Artificial Neural Networks*, *Support Vector Machines*, dan *Ensemble Learning*. Namun, hasil analisis menunjukkan bahwa performa model ML belum menunjukkan perbaikan yang signifikan dibandingkan dengan metode statistik tradisional. (Li et al., 2021)

Studi yang dilakukan oleh Ganggayah et al pada tahun 2019 tentang penggunaan teknik *Machine Learning* untuk membangun model prediksi terhadap kanker payudara mendapatkan kutipan dari beberapa artikel terbaru. Salah satunya adalah studi yang dilakukan oleh Li J pada tahun 2021 tentang peninjauan sistematis terhadap penelitian yang menggunakan *Machine Learning* untuk memprediksi tingkat kelangsungan hidup pasien kanker payudara selama lima tahun. Dalam penelitian tersebut, *Decision Trees* menjadi metode *Machine Learning* yang paling sering digunakan dengan akurasi model yang belum menunjukkan perbaikan yang signifikan dibandingkan dengan metode statistik tradisional.

Kemudian pada tahun 2022, Nik Ab Kadir MN et al. melakukan penelitian di Malaysia untuk mengembangkan model prediktif untuk bertahan hidup di kalangan wanita penderita kanker payudara dengan menggunakan Cox proportional hazard, jaringan saraf tiruan, dan analisis klasifikasi pohon keputusan. Dalam penelitian tersebut, model Cox PH memiliki akurasi tertinggi dibandingkan dengan model DT dan ANN dalam memprediksi kelangsungan hidup pasien.

4. KESIMPULAN

Analisis bibliometric dapat digunakan untuk menganalisis kinerja penulis dan pemetaan ilmiah dalam struktur publikasi ilmiah. Analisis kinerja penulis adalah mengevaluasi produktivitas dan dampak dari artikel menggunakan indikator bibliometrik. Demikian juga dengan pemetaan ilmiah dengan ditunjukkan struktur konseptual, struktur sosial, dan struktur kutipan menggunakan analisis co-occurrence pada keyword plus, analisis co-author, dan analisis co-reference. Untuk menganalisis *bibliometric* menggunakan paket Bibliometrix Rstudio. Menggunakan *Database Scopus* untuk koleksi literatur. Sebagai batasan menggunakan keywords dan tahun untuk membatasi jumlah literatur yang dianalisis. Penelitian ini menghasilkan temuan bahwa tiga negara yang menghasilkan artikel terbanyak adalah China, USA dan Iran. Sedangkan tiga negara yang menghasilkan sitasi terbanyak adalah China, USA dan India. Studi lebih lanjut dapat dilakukan dengan menambah koleksi literatur dan selektif terhadap *keyword* yang digunakan untuk mendapatkan temuan yang lebih komprehensif.

REFERENSI

- Abbas, S., Jalil, Z., Javed, A. R., Batool, I., Khan, M. Z., Noorwali, A., Gadekallu, T. R., & Akbar, A. (2021). BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized Tree algorithm. *PeerJ Computer Science*, 7, 1–20. <https://doi.org/10.7717/peerj-cs.390>
- Al-Azzam, N., & Shatnawi, I. (2021). Comparing supervised and semi-supervised *Machine Learning* Models on Diagnosing Breast Cancer. *Annals of Medicine and Surgery*, 62, 53–64. <https://doi.org/10.1016/j.amsu.2020.12.043>

- Alshammari, M., & Mezher, M. (2020). A comparative analysis of data mining techniques on breast cancer diagnosis data using WEKA toolbox. *International Journal of Advanced Computer Science and Applications*, 11(8), 224–229. <https://doi.org/10.14569/IJACSA.2020.0110829>
- Alromema, N., Syed, A. H., & Khan, T. (2023). A Hybrid Machine Learning Approach to Screen Optimal Predictors for the Classification of Primary Breast Tumors from Gene Expression Microarray Data. *Diagnostics*, 13(4), 708. <https://doi.org/10.3390/diagnostics13040708>
- Assegie, T. A., Tulasi, R. L., & Kumar, N. K. (2021). Breast cancer prediction model with Decision Tree and adaptive boosting. *IAES International Journal of Artificial Intelligence*, 10(1), 184–190. <https://doi.org/10.11591/ijai.v10.i1.pp184-190>
- Botlagunta, M., Botlagunta, M. D., Myneni, M. B., Lakshmi, D., Nayyar, A., Gullapalli, J. S., & Shah, M. A. (2023). Classification and diagnostic prediction of breast cancer metastasis on clinical data using Machine Learning algorithms. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-27548-w>
- Chen, L., Pan, X., Zhang, Y.-H., Hu, X., Feng, K., Huang, T., & Cai, Y.-D. (2019). Primary tumor site specificity is preserved in patient-derived tumor xenograft models. *Frontiers in Genetics*, 10(JUL). <https://doi.org/10.3389/fgene.2019.00738>
- Chen, X., Zhu, C.-C., & Yin, J. (2019). Ensemble of Decision Tree reveals potential miRNA-disease associations. *PLoS Computational Biology*, 15(7). Scopus. <https://doi.org/10.1371/journal.pcbi.1007209>
- Chowdhary, C. L., Mittal, M., Kumaresan, P., Pattanaik, P. A., & Marszalek, Z. (2020). An efficient segmentation and classification system in medical images using intuitionist possibilistic fuzzy C-mean clustering and fuzzy SVM algorithm. *Sensors (Switzerland)*, 20(14), 1–20. <https://doi.org/10.3390/s20143903>
- Dobrovaska, L., & Nosovets, O. (2021). Development Of The Classifier Based On A Multilayer Perceptron Using Genetic Algorithm And Cart Decision Tree. *Eastern-European Journal of Enterprise Technologies*, 5(9–113), 82–90. <https://doi.org/10.15587/1729-4061.2021.242795>
- El-Nabawy, A., Belal, N. A., & El-Bendary, N. (2021). A cascade deep forest model for breast cancer subtype classification using multi-omics data. *Mathematics*, 9(13). <https://doi.org/10.3390/math9131574>
- Feng, H., Wang, H., Xu, L., Ren, Y., Ni, Q., Yang, Z., Ma, S., Deng, Q., Chen, X., Xia, B., Kuang, Y., & Li, X. (2022). 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. *Frontiers in Oncology*, 12. <https://doi.org/10.3389/fonc.2022.1017435>
- Ferraro, D., Champ, J., Teste, B., Serra, M., Malaquin, L., Viovy, J.-L., de Cremoux, P., & Descroix, S. (2016). Microfluidic platform combining droplets and magnetic tweezers: Application to HER2 expression in cancer diagnosis. *Scientific Reports*, 6(1), 25540. <https://doi.org/10.1038/srep25540>
- Fusco, R., Di Marzo, M., Sansone, C., Sansone, M., & Petrillo, A. (2017). Breast DCE-MRI: lesion classification using dynamic and morphological features by means of a multiple classifier system. *European Radiology Experimental*, 1(1). <https://doi.org/10.1186/s41747-017-0007-4>
- Fusco, R., Granata, V., Raso, M. M., Vallone, P., De Rosa, A. P., Siani, C., Di Bonito, M., Petrillo, A., & Sansone, M. (2021). 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*, 13(10). <https://doi.org/10.3390/cancers13102421>
- Fusco, R., Piccirillo, A., Sansone, M., Granata, V., Vallone, P., Barretta, M. L., Petrosino, T., Siani, C., Di Giacomo, R., Di Bonito, M., Botti, G., & Petrillo, A. (2021). 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. *Applied Sciences (Switzerland)*, 11(4), 1–16. <https://doi.org/10.3390/app11041880>
- Ganggayah, M. D., Taib, N. A., Har, Y. C., Lio, P., & Dhillon, S. K. (2019). Predicting factors for survival of breast cancer patients using Machine Learning techniques. *BMC Medical Informatics and Decision Making*, 19(1). <https://doi.org/10.1186/s12911-019-0801-4>
- Gu, D., Zhao, W., Xie, Y., Wang, X., Su, K., & Zolotarev, O. V. (2021). A personalized medical decision support system based on explainable Machine Learning algorithms and ecc features: Data from the real world. *Diagnostics*, 11(9). <https://doi.org/10.3390/diagnostics11091677>
- Hadi, N. I., Jamal, Q., Iqbal, A., Shaikh, F., Somroo, S., & Musharraf, S. G. (2017). Serum Metabolomic Profiles for Breast Cancer Diagnosis, Grading and Staging by Gas Chromatography-Mass Spectrometry. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-01924-9>
- Huang, H., Feng, X., Zhou, S., Jiang, J., Chen, H., Li, Y., & Li, C. (2019). A new fruit fly optimization algorithm enhanced support vector machine for diagnosis of breast cancer based on high-level features. *BMC Bioinformatics*, 20. <https://doi.org/10.1186/s12859-019-2771-z>
- Huang, Z., & Chen, D. (2022). A Breast Cancer Diagnosis Method Based on VIM Feature Selection and Hierarchical Clustering Random Forest Algorithm. *IEEE Access*, 10, 3284–3293. <https://doi.org/10.1109/ACCESS.2021.3139595>
- Hussain, L., Huang, P., Nguyen, T., Lone, K. J., Ali, A., Khan, M. S., Li, H., Suh, D. Y., & Duong, T. Q. (2021). Machine Learning classification of texture features of MRI breast tumor and peritumor of combined pre- and early treatment predicts pathologic complete response. *BioMedical Engineering Online*, 20(1). <https://doi.org/10.1186/s12938-021-00899-z>
- Jayatilake, S. M. D. A. C., & Ganegoda, G. U. (2021). Involvement of Machine Learning Tools in Healthcare Decision Making. *Journal of Healthcare Engineering*, 2021. <https://doi.org/10.1155/2021/6679512>
- Jiang, N., Tian, T., Chen, X., Zhang, G., Pan, L., Yan, C., Yang, G., Wang, L., Cao, X., & Wang, X. (2021). A Diagnostic Analysis Workflow to Optimal Multiple Tumor Markers to Predict the Nonmetastatic Breast Cancer from Breast Lumps. *Journal of Oncology*, 2021. <https://doi.org/10.1155/2021/5579373>
- Kaur, P., Singh, G., & Kaur, P. (2019). Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. *Informatics in Medicine Unlocked*, 16. <https://doi.org/10.1016/j.imu.2019.01.001>
- Kaushik, A. C., Mehmood, A., Wang, X., Wei, D.-Q., & Dai, X. (2021). Globally ncRNAs Expression Profiling of TNBC and Screening of Functional lncRNA. *Frontiers in Bioengineering and Biotechnology*, 8. <https://doi.org/10.3389/fbioe.2020.523127>
- Kaya Keleş, M. (2019). Breast cancer prediction and detection using data mining classification algorithms: A comparative study. *Tehnicki Vjesnik*, 26(1), 149–155. Scopus. <https://doi.org/10.17559/TV-20180417102943>
- Khan, M. B. S., Atta-Ur-Rahman, Nawaz, M. S., Ahmed, R., Khan, M. A., & Mosavi, A. (2022). Intelligent breast cancer diagnostic system empowered by deep extreme gradient descent optimization. *Mathematical Biosciences and Engineering*, 19(8), 7978–8002. <https://doi.org/10.3934/mbe.2022373>
- Lan, X., Wang, X., Qi, J., Chen, H., Zeng, X., Shi, J., Liu, D., Shen, H., & Zhang, J. (2022). Application of Machine Learning with multiparametric dual-energy computed tomography of the breast to differentiate between benign and malignant lesions. *Quantitative Imaging in Medicine and Surgery*, 12(1), 810–822. <https://doi.org/10.21037/qims-21-39>

- Lei, Y.-M., Yin, M., Yu, M.-H., Yu, J., Zeng, S.-E., Lv, W.-Z., Li, J., Ye, H.-R., Cui, X.-W., & Dietrich, C. F. (2021). Artificial Intelligence in Medical Imaging of the Breast. *Frontiers in Oncology*, 11. <https://doi.org/10.3389/fonc.2021.600557>
- Liang, H., Li, J., Wu, H., Li, L., Zhou, X., & Jiang, X. (2022). Mammographic Classification of Breast Cancer Microcalcifications through Extreme Gradient Boosting. *Electronics (Switzerland)*, 11(15). <https://doi.org/10.3390/electronics11152435>
- Li, G., Fang, T., Zhang, Y., Liang, C., Xiao, Q., & Luo, J. (2022). Predicting miRNA-disease associations based on graph attention network with multi-source information. *BMC Bioinformatics*, 23(1). <https://doi.org/10.1186/s12859-022-04796-7>
- Li, J., Zhou, Z., Dong, J., Fu, Y., Li, Y., Luan, Z., & Peng, X. (2021). Predicting breast cancer 5-year survival using Machine Learning: A systematic review. *PLoS ONE*, 16. <https://doi.org/10.1371/journal.pone.0250370>
- Massafra, R., Latorre, A., Fanizzi, A., Bellotti, R., Didonna, V., Giotta, F., La Forgia, D., Nardone, A., Pastena, M., Ressa, C. M., Rinaldi, L., Russo, A. O. M., Tamborra, P., Tangaro, S., Zito, A., & Lorusso, V. (2021). A Clinical Decision Support System for Predicting Invasive Breast Cancer Recurrence: Preliminary Results. *Frontiers in Oncology*, 11. <https://doi.org/10.3389/fonc.2021.576007>
- Min, Y., Wei, X., Chen, H., Xiang, K., Yin, G., & Feng, Y. (2021). Identifying Clinicopathological Risk Factors of the Regional Lymph Node Metastasis in Patients with T1-2Mucinous Breast Cancer: A Population-Based Study. *Journal of Oncology*, 2021. <https://doi.org/10.1155/2021/3866907>
- Mohebian, M. R., Marateb, H. R., Mansourian, M., Mañanas, M. A., & Mokarian, F. (2017). A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning. *Computational and Structural Biotechnology Journal*, 15, 75–85. <https://doi.org/10.1016/j.csbj.2016.11.004>
- Mudunuru, V. R., & Skrzypek, L. A. (2020). A comparison of artificial neural network and Decision Trees with logistic regression as classification models for breast cancer survival. *International Journal of Mathematical, Engineering and Management Sciences*, 5(6), 1170–1190. <https://doi.org/10.33889/IJMEMS.2020.5.6.089>
- Nahid, A.-A., & Kong, Y. (2017). Involvement of Machine Learning for Breast Cancer Image Classification: A Survey. *Computational and Mathematical Methods in Medicine*, 2017. Scopus. <https://doi.org/10.1155/2017/3781951>
- Nasser, F. K., & Behadili, S. F. (2022). Breast Cancer Detection using *Decision Tree* and K-Nearest Neighbour Classifiers. *Iraqi Journal of Science*, 63(11), 4987–5003. <https://doi.org/10.24996/ij.s.2022.63.11.34>
- Nik Ab Kadir, M. N., Yaacob, N. M., Yusof, S. N., Ab Hadi, I. S., Musa, K. I., Mohd Isa, S. A., Bahtiar, B., Adam, F., Yahya, M. M., & Hairon, S. M. (2022). Development of Predictive Models for Survival among Women with Breast Cancer in Malaysia. *International Journal of Environmental Research and Public Health*, 19(22). <https://doi.org/10.3390/ijerph192215335>
- Osman, A. H., & Aljahdali, H. M. A. (2020). An Effective of Ensemble Boosting Learning Method for Breast Cancer Virtual Screening Using Neural Network Model. *IEEE Access*, 8, 39165–39174. <https://doi.org/10.1109/ACCESS.2020.2976149>
- Ozcan, I., Aydin, H., & Cetinkaya, A. (2022). Comparison of Classification Success Rates of Different Machine Learning Algorithms in the Diagnosis of Breast Cancer. *Asian Pacific Journal of Cancer Prevention*, 23(10), 3287–3297. <https://doi.org/10.31557/APJCP.2022.23.10.3287>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M.,

- Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 89. <https://doi.org/10.1186/s13643-021-01626-4>
- Pelon, F., Bourachot, B., Kieffer, Y., Magagna, I., Mermet-Meillon, F., Bonnet, I., Costa, A., Givel, A.-M., Attieh, Y., Barbazan, J., Bonneau, C., Fuhrmann, L., Descroix, S., Vignjevic, D., Silberzan, P., Parrini, M. C., Vincent-Salomon, A., & Mechta-Grigoriou, F. (2020). Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-019-14134-w>
- Peng, Y., Li, W., & Liu, Y. (2006). A Hybrid Approach for Biomarker Discovery from Microarray Gene Expression Data for Cancer Classification. *Cancer Informatics*, 2, 117693510600200. <https://doi.org/10.1177/117693510600200024>
- Qawqzeh, Y. K., Alourani, A., & Ghwanmeh, S. (2023). An Improved Breast Cancer Classification Method Using an Enhanced AdaBoost Classifier. *International Journal of Advanced Computer Science and Applications*, 14(1), 473–478. <https://doi.org/10.14569/IJACSA.2023.0140151>
- Rexhepaj, E., Jirstrom, K., O'Connor, D. P., O'Brien, S. L., Landberg, G., Duffy, M. J., Brennan, D. J., & Gallagher, W. M. (2010). Validation of cytoplasmic-to-nuclear ratio of survivin as an indicator of improved prognosis in breast cancer. *BMC Cancer*, 10(1), 639. <https://doi.org/10.1186/1471-2407-10-639>
- Saarela, M., & Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences*, 3(2). <https://doi.org/10.1007/s42452-021-04148-9>
- Sakri, S. B., Abdul Rashid, N. B., & Muhammad Zain, Z. (2018). Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction. *IEEE Access*, 6, 29637–29647. <https://doi.org/10.1109/ACCESS.2018.2843443>
- Sánchez-Calderón, D., Pedraza, A., Urrego, C. M., Mejía-Mejía, A., Montealegre-Páez, A. L., & Perdomo, S. (2020). Analysis of the cost-effectiveness of liquid biopsy to determine treatment change in patients with her2-positive advanced breast cancer in Colombia. *ClinicoEconomics and Outcomes Research*, 12, 115–122. <https://doi.org/10.2147/CEOR.S220726>
- Sansone, M., Fusco, R., Grassi, F., Gatta, G., Belfiore, M. P., Angelone, F., Ricciardi, C., Ponsiglione, A. M., Amato, F., Galdiero, R., Grassi, R., Granata, V., & Grassi, R. (2023). Machine Learning Approaches with Textural Features to Calculate Breast Density on Mammography. *Current Oncology*, 30(1), 839–853. <https://doi.org/10.3390/curroncol30010064>
- Shanbehzadeh, M., Kazemi-Arpanahi, H., Bolbolian Ghalibaf, M., & Orooji, A. (2022). Performance evaluation of Machine Learning for breast cancer diagnosis: A case study. *Informatics in Medicine Unlocked*, 31. <https://doi.org/10.1016/j.imu.2022.101009>
- Smerekanych, S., Johnson, T. S., Huang, K., & Zhang, Y. (2020). Pseudogene-gene functional networks are prognostic of patient survival in breast cancer. *BMC Medical Genomics*, 13. Scopus. <https://doi.org/10.1186/s12920-020-0687-0>
- Su, F., Gao, Z., Liu, Y., Zhou, G., Cui, Y., Deng, C., Liu, Y., Zhang, Y., Ma, X., Wang, Y., Guan, L., Zhang, Y., & Liu, B. (2022). 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*, 13(11). <https://doi.org/10.3390/genes13111931>
- Sun, L., He, J., Yin, X., Zhang, Y., Chen, J.-H., Kron, T., & Su, M.-Y. (2018). An image segmentation framework for extracting tumors from breast magnetic resonance images. *Journal of Innovative Optical Health Sciences*, 11(4). <https://doi.org/10.1142/S1793545818500141>

- Thareja, P., & Chhillar, R. S. (2021). Comparative Analysis of Data Mining Algorithms for Cancer Gene Expression Data. *International Journal of Advanced Computer Science and Applications*, 12(10), 322–328. <https://doi.org/10.14569/IJACSA.2021.0121035>
- Tian, J.-X., & Zhang, J. (2022). Breast cancer diagnosis using feature extraction and boosted C5.0 Decision Tree algorithm with penalty factor. *Mathematical Biosciences and Engineering*, 19(3), 2193–2205. <https://doi.org/10.3934/MBE.2022102>
- Wang, K., Li, L., Franch-Expósito, S., Le, X., Tang, J., Li, Q., Wu, Q., Bassaganyas, L., Camps, J., Zhang, X., Li, H., Foukakis, T., Xiang, T., Wu, J., & Ren, G. (2022). Integrated multi-omics profiling of high-grade estrogen receptor-positive, HER2-negative breast cancer. *Molecular Oncology*, 16(12), 2413–2431. <https://doi.org/10.1002/1878-0261.13043>
- Xie, T., Wang, Z., Zhao, Q., Bai, Q., Zhou, X., Gu, Y., Peng, W., & Wang, H. (2019). Machine Learning-based analysis of MR multiparametric radiomics for the subtype classification of breast cancer. *Frontiers in Oncology*, 9(JUN). <https://doi.org/10.3389/fonc.2019.00505>
- XI, G., HE, J., KANG, D., XU, S., GUO, W., FU, F., LIU, Y., ZHENG, L., QIU, L., LI, L., WANG, C., & CHEN, J. (2021). Nomogram model combining macro and micro tumor-associated collagen signatures obtained from multiphoton images to predict the histologic grade in breast cancer. *Biomedical Optics Express*, 12(10), 6558–6570. <https://doi.org/10.1364/BOE.433281>
- Xiong, F., Cao, X., Shi, X., Long, Z., Liu, Y., & Lei, M. (2022). A Machine Learning–Based model to predict early death among bone metastatic breast cancer patients: A large cohort of 16,189 patients. *Frontiers in Cell and Developmental Biology*, 10. <https://doi.org/10.3389/fcell.2022.1059597>
- Xu, J., Rao, X., Lu, W., Xie, X., Wang, X., & Li, X. (2022). Noninvasive Predictor for Premalignant and Cancerous Lesions in Endometrial Polyps Diagnosed by Ultrasound. *Frontiers in Oncology*, 11. <https://doi.org/10.3389/fonc.2021.812033>
- Yang, L., Chang, J., He, X., Peng, M., Zhang, Y., Wu, T., Xu, P., Chu, W., Gao, C., Cao, S., & Kang, S. (2022). PET/CT-based radiomics analysis may help to predict neoadjuvant chemotherapy outcomes in breast cancer. *Frontiers in Oncology*, 12. <https://doi.org/10.3389/fonc.2022.849626>
- Yang, L., Zhang, Z., Li, J., Chen, M., Yang, J., Fu, J., Bu, H., Tang, S., Liu, Y., Li, H., Li, X., Xu, F., Teng, X., Yang, Y., Ma, Y., Guo, S., Wang, J., & Guo, D. (2018). 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. *Journal of Cancer*, 9(13), 2327–2333. <https://doi.org/10.7150/jca.25586>
- Yue, W., Wang, Z., Chen, H., Payne, A., & Liu, X. (2018). Machine Learning with applications in breast cancer diagnosis and prognosis. *Designs*, 2(2), 1–17. <https://doi.org/10.3390/designs2020013>
- Zhao, L., Xie, S., Zhou, B., Shen, C., Li, L., Pi, W., Gong, Z., Zhao, J., Peng, Q., Zhou, J., Peng, J., Zhou, Y., Zou, L., Song, L., Zhu, H., & Luo, H. (2022). Machine Learning Algorithms Identify Clinical Subtypes and Cancer in Anti-TIF1γ+ Myositis: A Longitudinal Study of 87 Patients. *Frontiers in Immunology*, 13. <https://doi.org/10.3389/fimmu.2022.802499>
- Zhang, Y., Li, J., Fan, Y., Li, X., Qiu, J., Zhu, M., & Li, H. (2019). Risk factors for axillary lymph node metastases in clinical stage T1-2N0M0 breast cancer patients. *Medicine (United States)*, 98(40). <https://doi.org/10.1097/MD.00000000000017481>
- Zhang, Y., Zhou, Y., Mao, F., Yao, R., & Sun, Q. (2020). Ki-67 index, progesterone receptor expression, histologic grade and tumor size in predicting breast cancer recurrence risk: A consecutive cohort study. *Cancer Communications*, 40(4), 181–193. <https://doi.org/10.1002/cac2.12024>