Environmental Science and Pollution Research (2023) 30:96125–96137
https://doi.org/10.1007/s11356-023-28576-9

RESEARCH ARTICLE

Bibliometric analysis of the global scientific production on machine learning applied to different cancer types

Miguel Angel Ruiz-Fresneda[1] · Alfonso Gijón[2,3] · Pablo Morales-Álvarez[3,4]

Received: 27 January 2023 / Accepted: 29 June 2023 / Published online: 11 August 2023
© The Author(s) 2023

Abstract

Cancer disease is one of the main causes of death in the world, with million annual cases in the last decades. The need to find a cure has stimulated the search for efficient treatments and diagnostic procedures. One of the most promising tools that has emerged against cancer in recent years is machine learning (ML), which has raised a huge number of scientific papers published in a relatively short period of time. The present study analyzes global scientific production on ML applied to the most relevant cancer types through various bibliometric indicators. We find that over 30,000 studies have been published so far and observe that cancers with the highest number of published studies using ML (breast, lung, and colon cancer) are those with the highest incidence, being the USA and China the main scientific producers on the subject. Interestingly, the role of China and Japan in stomach cancer is correlated with the number of cases of this cancer type in Asia (78% of the worldwide cases). Knowing the countries and institutions that most study each area can be of great help for improving international collaborations between research groups and countries. Our analysis shows that medical and computer science journals lead the number of publications on the subject and could be useful for researchers in the field. Finally, keyword co-occurrence analysis suggests that ML-cancer research trends are focused not only on the use of ML as an effective diagnostic method, but also for the improvement of radiotherapy- and chemotherapy-based treatments.

Keywords Machine learning · Cancer · Bibliometric analysis · Artificial intelligence · Public health

Introduction

The thriving development in sensor and storage technology has enabled the collection of ever-increasing amounts of data (Shilo et al. 2020; Nathan et al. 2022). This growth in data availability is affecting many different fields of research, well beyond computer science and engineering. For example, the increasing digitalization of medical tests such as X-rays, biopsies, electrocardiograms, and blood tests is laying the foundations for personalized medicine (Vokinger and Gasser 2021; Houssein et al. 2021). Likewise, thousands of satellites daily provide unprecedented data streams as remote sensing images for earth-observation and climate change analysis. Also, the evaluations provided by customers in platforms such as Amazon or Netflix generate plenty of relevant information for marketing campaigns.

All this raw data is not enough on its own to make decisions. Indeed, the data must be analyzed by an expert, i.e., a clinician, an environmental scientist, or a business administrator in the three examples above, respectively. Due to the huge amount of available data, this is a daunting task for a single expert, or even for a team of them. The goal of artificial intelligence (AI) is to automate this process of knowledge extraction from data. Early approaches in AI were based on expert systems (Shortliffe 1986; Duan et al. 2005).

Responsible editor: Philippe Garrigues

✉ Miguel Angel Ruiz-Fresneda
mafres@ugr.es

Alfonso Gijón
alfonso.gijon@ugr.es

Pablo Morales-Álvarez
pablomorales@ugr.es

[1] Department of Microbiology, University of Granada, Granada, Spain

[2] Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain

[3] Research Centre for Information and Communication Technologies (CITIC-UGR), University of Granada, Granada, Spain

[4] Department of Statistics and Operations Research, University of Granada, Granada, Spain

The idea in expert systems is to hard-code the knowledge of the expert in some formal language so that the machine can apply it. Although this paradigm has proved useful for very structured tasks, it struggles in other problems such as general object and speech recognition, which require subjective and subtle knowledge that cannot be easily codified in formal computer languages (Saibene et al. 2021).

Machine learning (ML) has emerged as a different paradigm to extract knowledge from data (Ravuri et al. 2018; Hameed et al. 2020). Instead of hard-coding the knowledge of an expert, ML algorithms try to learn their own knowledge based on specific examples of the task at hand (Murphy 2012; López-Pérez et al. 2021). This has led to much more accurate results, since ML algorithms learn to extract relevant features and are able to reason based on them (Goodfellow et al. 2016). For instance, consider the problem of cancer detection in histopathological images. An expert-system tries to codify the knowledge of a pathologist into a sequence of instructions that can be systematically applied by a computer (e.g., looking for regions with a certain color or shape). Alternatively, ML algorithms are shown several examples of (labeled) cancerous and non-cancerous images, and the algorithm learns its own rules to make predictions (López-Úbeda et al. 2020).

Many different algorithms have been developed in the machine learning community. One of the most popular approaches is given by artificial neural networks, also known as deep learning, which leverage several layers of simple operations to extract increasingly abstract features that can be used for reasoning (LeCun et al. 2015). For example, convolutional neural networks have achieved astonishing results in image processing, and recurrent neural networks have excelled at speech recognition. Another important family of algorithms is given by probabilistic kernel methods, such as support vector machines (SVM) (Akay 2009; Chen et al. 2011) and Gaussian processes (GP) (Wang et al. 2019; Morales-Álvarez et al. 2022). The latter has become increasingly popular due to its capability to quantify uncertainty, which is essential for real-world applications of machine learning.

Although machine learning has been used in many different applications, in this paper, we focus on the medical domain. More specifically, we are interested in the problem of cancer, which has been studied through machine learning from different perspectives. For example, digital or computational pathology leverages ML algorithms to detect the presence of cancer in digitalized biopsies (Peng et al. 2022; Nguyen et al. 2022; Ain et al. 2022). The goal here is to speed up the cancer detection process, to ensure the democratization of early cancer diagnosis. Machine learning is also used for basic cancer research, in order to analyze the properties of molecules and drugs that can lead to potential treatments (Vamathevan et al. 2019). Likewise, ML algorithms are deployed to improve the treatment of oncology patients, analyzing the results of tumor markers throughout the radiotherapy and chemotherapy processes (Cuocolo et al. 2020). It is important to stress out that cancer is one of the main challenges for the XXI century, as it is the second leading cause of death worldwide according to the American Cancer Society (10 million deaths in 2020 were attributed to cancer) (American Cancer Society 2021).

In order to evaluate and optimize the current huge investment in ML for cancer research, the total volume of scientific production in the field must be analyzed. Bibliometric data analyses can be very useful in the understanding and classification of such a large amount of published documents and can shed light on the development of both ongoing and new research. The aim of the present study is to analyze global scientific production, impact, and research trends on ML applied to the types of cancer that present the highest incidence (in terms of death rate). Many previous bibliometric reviews have focused on specific cancers types, such as breast cancer (Salod and Singh 2020; Joshi et al. 2021), rectal and colorectal tumors (Wang et al. 2020; Kennion et al. 2022), or gynecological ones (Fiste et al. 2022). Whereas these works cover individual cancer types in depth, they do not provide a global unified bibliometric analysis of the most frequent ones. Other works have focused on literature related to specific stages of cancer disease, regardless of the cancer type, such as cancer rehabilitation (Tschuggnall et al. 2021) and cancer survival prediction (Deepa and Gunavathi 2022). There also exist insightful reviews on the most popular ML techniques for cancer research (see Maurya et al. 2023; Mokoatle et al. 2023). But notice that these works focus on the methodological aspects of the ML approaches and do not provide a bibliometric perspective of the field. In contrast to previous work, here we present a unified, novel, updated, and comparative quantitative study for each one of the most important cancers in the last years according to the World Health Organization (lung, colorectal, liver, stomach, and breast cancer). The results presented here are expected to encourage international collaborations between countries and research institutions and to favor the development of new research in the field.

The rest of this paper is organized as follows: The "Methods" section introduces the research methodology, including the search strategies as well as the data processing and analysis. The "Results and discussion" section presents and discusses the main results on the scientific production of machine learning applied to cancer. Finally, the "Conclusions" section summarizes the main conclusions of this work.

Methods

We have gathered our data from scientific production indexed in the Web of Science core collection databases

(WOS 2022). This multi-disciplinary international source references the most prestigious scientific publications in the world and is an essential starting point for bibliometric studies providing indicators of production and scientific impact. We launched our searches from 1900 to 31-12-2021 comprising almost all year's timespan. The search flow is shown in Fig. 1, where different combinations and number of documents are included.

Search strategies

To gather data comparing scientific production on machine learning for the study of different types of cancer, we conducted searches in WOS "Web of science core collection">*Advanced Search*>*TS*=*Topic*, as summarized in Fig. 1. Topic search strategy includes title, abstract, author keywords, and keywords plus.

WEB OF SCIENCE - ADVANCED SEARCH	
A) DOCUMENTS	B) RESEARCHERS
1. SELECTING DATABASE Search in: WoS Core Collection Editions: all	
2. DEFINING SEACH STRATEGY	
TOPIC (TS) + TERMS	**DOCUMENTS**
SEARCH #1	809.708
SEARCH #2	4.306.371
SEARCH #3	1.296.840
SEARCH #4	458.607
SEARCH #5	1.119.063
SEARCH #6	388.946
SEARCH #7	779.740
3. DATA PROCESSING AND ANALYSIS	
COMBINATIONS	**DOCUMENTS**
COMBINE #1 AND #2	31.169
COMBINE #1 AND #2 AND #3	3.997
COMBINE #1 AND #2 AND #4	2.135
COMBINE #1 AND #2 AND #5	1.730
COMBINE #1 AND #2 AND #6	736
COMBINE #1 AND #2 AND #7	7.288

Fig. 1 Flow diagram summarizing the search strategy and analysis performed using the Web of Science (WoS)

Firstly, we designed a general search of machine learning research on cancer over time. For this purpose, the #1 search was performed to identify documents that studied machine learning in general terms by using the equation: *TS*=("Machine Learning" *OR* "Data Science" *OR* "Machine Intelligence" *OR* "Data mining" *OR* "Big data" *OR* "Artificial Intelligence" *OR* "Deep Learning" *OR* "Deep Learn" *OR* "Supervised Learning" *OR* "Unsupervised Learning" *OR* "Neural networks" *OR* "Convolutional Neural Network" *OR* "Reinforcement Learning" *OR* "Natural Language Processing" *OR* "Natural Language Process" *OR* "Artificial neural network"). *TimeSpan*=1900-2021. After that, the #2 search was performed using the equation: *TS*=("Cancer*" *OR* "tumor*" *OR* "neoplasia*" *OR* "neoplasm*" *OR* "oncology" *OR* "metastasis" *OR* "metastatic" *OR* "carcinoma"). *TimeSpan*=1900-2021. To identify within the set of documents retrieved in #1, those that studied cancer, we constructed the intersection between the search strategies #1 and #2, by using "Combine #1 *AND* #2".

Subsequently, we designed a search comparing the use of machine learning for certain types of cancer. Specifically, we focused on types of cancer that caused the highest number of death in 2020, according to the World health Organization (WHO): lung cancer (1.8 million deaths), colon and rectum cancer (916,000 deaths), liver cancer (830,000 deaths), stomach cancer (769,000 deaths), and breast cancer (685,000 deaths). As a results, the #3 search included a list of terms about lung cancer: *TS*=("lung" *OR* "pulmonary" *OR* "pulmonic"). *TimeSpan*=1900-2021. The search #4 was constructed for colon and rectum cancer: *TS*=("colon" *OR* "rectum" *OR* "colorectal" *OR* "large intestine"). *TimeSpan*=1900-2021. The #5 search included liver cancer terms: *TS*=("liver" *OR* "hepatocellular" *OR* "hepatoma"). *TimeSpan*=1900-2021. Finally, the #6 search was constructed for stomach cancer: *TS*=("stomach" *OR* "gastric") (*TimeSpan*=1900-2021), and the #7 search for breast cancer: *TS*=("breast"). *TimeSpan*=1900-2021.

Data processing and analysis

Data obtained from the search "Combine #1 *AND* #2" were tabulated, and we produced a table of annual scientific production on machine learning applied to cancer studies by institution, country, and journal. The 31,169 reported documents resulting from the search "combine #1 *AND* #2" were processed and standardized in Excel.

For the individual analysis of each cancer type resulting from the searches "combine #1 *AND* #2 *AND* #3," "combine #1 *AND* #2 *AND* #4," "combine #1 *AND* #2 *AND* #5," "combine #1 *AND* #2 *AND* #6," "combine #1 *AND* #2 *AND* #7,"

we designed a database to analyze the production and impact of the studies recorded about machine learning, considering TSP (total studied produced), CR (citations received), MCS (mean citations/study), CS (citing studies), +CS (citations received by the most cited work), and H-index (number of studies that have received the same or a higher number of citations). Additionally, we designed a database to analyze the top 5 production of the studies recorded disseminated by institutions, producer countries, and journals. Finally, visualization network mapping for co-occurrence keywords was performed for each cancer type to analyze the global trends on the topic. To visualize the bibliometric networks, we used VOS-viewer software (https://www.vosviewer.com/), which works with units of analysis (authors, organizations, keywords, etc.) and of measurement (links, frequency, centrality, distance), to illustrate our results by grouping similarities in clusters. To build the co-occurrence networks, we generated vectors, which were pre-displayed in PAJEK (http://mrvar.fdv.uni-lj.si/pajek/), with definitive drawings created in VOS-viewer. We used this process because VOS-viewer is limited in that it labels nodes based on an internal, non-modifiable schedule. We labeled as many nodes as possible while guaranteeing the set were correctly displayed.

Results and discussion

In this section, we present and discuss our main results. We first study the more general field of machine learning applied to any cancer type (see "Overview of scientific production on machine learning applied to cancer"). Then, we separately focus on the five cancer types with highest incidence (see "Comparison of scientific production on different types of cancers" and "Analysis of keywords co-occurrence on different types of cancers"). Finally, based on the literature identified in this section, we briefly discuss the potential of machine learning in cancer research (see "Machine learning in cancer research: a paradigm shift in diagnosis, treatment, and beyond").

Overview of scientific production on machine learning applied to cancer

Our search showed a high amount of studies published on machine learning applied to cancer research until 2021, with a total of 31,169 documents (see Table 1). The first study applying ML to cancer was published in 1983 as a meeting abstract in the journal Medical Physics, with the title "An artificial-intelligence program to plan radiotherapy for cancer of the oral cavity" (Paluszynski et al. 1983). However, a solid interest was not observed until 3 decades later, when an exponential increase in the number of publications occurred in the decade of 2010. In fact, 75% of the documents (23,517 out of 31,169) have been published in the last 5 years analyzed (2017–2021). This may be due to the rapid development and advancement of ML in recent years in both computational resources and algorithms (Jordan and Mitchell 2015), as suggested in Fig. 2, where an increasing number of publications appear within that period. Besides, in the inset of the same figure, we clearly observe a correlation between the trends in the scientific production on ML only and ML applied to cancer, which indicates that ML techniques have been applied to cancer since their early days. Until 2017, the production in ML (green line) was approximately equal to 30 times the production in Cancer+ML (blue line), while from 2018 onward, the latter topic has grown even more rapidly in relative terms than general ML.

In turn, the need to find new methods to improve the diagnosis and treatment of a disease as serious as cancer (Jemal et al. 2011) has triggered a growing number of publications since the last century (see Fig. 2). The relevance of cancer

Table 1 TSP (total studies produced) per year on ML applied to cancer since 1983 until 2021

Year	TSP	Year	TSP	Year	TSP
1983	1	2000	120	2021	460
1988	4	2001	118	2013	557
1989	1	2002	149	2014	709
1991	12	2003	211	2015	888
1992	23	2004	227	2016	1180
1993	16	2005	276	2017	1770
1994	36	2006	320	2018	2895
1995	40	2007	376	2019	4750
1996	52	2008	355	2020	6342
1997	97	2009	414	2021	7760
1998	112	2010	388		
1999	89	2011	421	Total	31,169

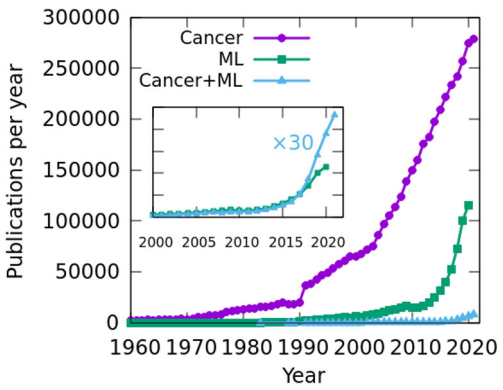

Fig. 2 Number of publications per year on cancer (violet), ML (green), and ML applied to cancer (blue), over time. The inset compares the ML and Cancer+ML cases, where the latter has been linearly scaled by a factor of 30 for a better visualization

has also motivated the application of ML techniques in that field as a useful tool. The importance of ML research applied to cancer can be appreciated by comparing the total scientific production that we have detected in the present study compared to the use of ML in other fields. Bibliometric studies analyzing applications of ML in physiological signals (Faust et al. 2018), sustainable manufacturing (Jamwal et al. 2021), Industry 4.0 (Muhuri et al. 2019), public health problems (dos Santos et al. 2019), maritime industry (Munim et al. 2020), management of depressive disorders (Tran et al. 2019), cybernetics (Xu et al. 2019), or operations environment (Dhamija and Bag 2020) found a total of 53, 96, 194, 250, 279, 544, and 1854 publications, respectively, far from the huge amount of 31,169 documents analyzed in the present study. Notice that the lower number of articles found by these authors could be also influenced by the use of a different search strategy, including different selection filters, search periods, types of documents, and database collection, for example.

Regarding the nationality of the institutional affiliation, US institutions are the leaders of the field with 4 institutions in the top 5: Harvard University (926 studies), University of California System (885 studies), University of Texas System (860 studies), and Harvard Medical School (501 studies), as can be seen in Table 2. Only one from China (The Chinese Academy of Science) is among the top five (fourth position) with 619 studies. This fact indicates the clear dominance and potential of American institutions on the investigated subject. However, it can be observed that the production is very distributed, since the institution with the most publications on the subject (Harvard University) accounts for 2.97% of the total number of published works (see Table 2), and the rest of the institutions present values that are not very different. According to these results, it is logical that in the ranking by countries, the leader is the USA, with a total of 10,051, which represents a huge amount of 32.25% of the total production. China follows with a total of 6721. India also stands out in this sector with 2375 documents (7.30%), which can be related to the great development gained in research in the field of medicine and pharmacology in recent years (Meena and Mathaiyan 2021). In terms of journals, the production is very distributed since none in particular includes the majority of studies. The one that published the most papers on the subject is Proceedings of Spie, with 826 and a total of 2.65% of the production. It should be noted that the 3 most producing journals are in the field of computer science, while the rest of the top 5 are journals dedicated to medicine.

Comparison of scientific production on different types of cancers

Following the analysis of the global scientific production on machine learning applied to cancer addressed in the last section, a specific and individualized study was carried out on different types of cancer with high incidence and mortality rates (lung, colon, liver, stomach, and breast cancer). The aim of such an analysis was to elucidate for the first time the differences or similarities, in terms of scientific production, between different types of cancer in order to obtain potentially relevant conclusions. Specifically, we analyze the scientific impact and production at the level of institutions, countries, and scientific journals.

Our results analyzing five types of cancer, shown in Table 3, indicated that breast cancer has been the most studied with ML techniques, with 7288 studies, followed by lung, colon, liver, and stomach cancer with 3997, 2135, 1730, and 736, respectively. Studies on breast cancer also have the highest impact with 128,171 total citations and an H-index of 137. Out of the 84,637 citing studies, the one that has received the most citations contains a total of 3826 total citations. Lung cancer is in second position in terms of scientific impact with a total of 69,640 citations, which supposes almost half of those received for breast cancer. As can be observed, colon and liver cancer present intermediate levels of impact in third (30,392) and fourth position (23,416), respectively. In last place is stomach cancer with 8662 citations received, well below the rest of cancer types.

The total production of studies (TSP) for each type of cancer (see Table 3) matches well with the data on articles with the highest number of citations. Indeed, the works by Ehteshami Bejnordi et al. (2017); Coudray et al. (2018); Ye et al. (2003); Sirinukunwattana et al. (2016); Kather et al. (2019) stand out in the fields of breast, lung, liver, colon/rectum, and gastric cancer with 1291, 1069, 684, 661, and 437 cita-

Table 2 TSP (total studies produced) and TSP percentage of the total of 31,169 studies produced, classified by institution, country, and journal

Institution	TSP	%TSP	Country	TSP	%TSP	Journal	TSP	%TSP
Harvard University	926	2.97	USA	10,051	32.25	Proceedings of Spie	826	2.65
University of California System	885	2.84	China	6271	21.56	Lecture Notes in Computer Science	668	2.14
University of Texas System	860	2.76	India	2275	7.30	Medical Physics	572	1.84
Chinese Academy of Sciences	619	1.99	England	2015	6.47	Scientific Reports	552	1.77
Harvard Medical School	501	1.61	Germany	1744	5.60	IEEE Access	445	1.43

Table 3 Bibliometric parameters classified by cancer type

Cancer type	TSP	CR	MCS	CS	+CR	H-index
Lung	3997	69,640	17.42	48,651	3826	110
Colon and rectum	2135	30,392	14.24	22,893	1181	76
Liver	1730	23,416	13.54	19,308	679	62
Stomach	736	8662	11.77	6515	261	43
Breast	7288	128,171	19.59	84,637	3826	137

TSP total studies produced, *CR* citations received, *MCS* mean citations/study, *CS* citing studies, *+CR* citations received by the most cited work, *H-index* number of studies that have received the same or a higher number of citations

tions, respectively. Based on these highly cited studies and some recent review publications, one can briefly analyze the most popular approaches in each of the explored areas. In the context of lung cancer, popular machine learning methods include convolutional neural networks (CNNs), support vector machines (SVMs), and random forests, which have been employed for tasks such as tumor classification, early detection, and survival prediction using medical images as inputs (Coudray et al. 2018; Li et al. 2022; Wang 2022). The same widely used supervised classifiers are leveraged for breast cancer, sometimes combined with long-short term memory (LSTM) networks (Ehteshami Bejnordi et al. 2017; Zhang et al. 2020). For colorectal cancer, on the one hand, we have studies using CNNs from images (Yu and Helwig 2021; Sirinukunwattana et al. 2016), and on the other hand, we have classification algorithms like random forest to analyze genetic and molecular data, aiding in predicting disease progression and survival outcome (Koppad et al. 2022). Similarly to colorectal cancer, in liver cancer research, there are applications of CNNs to medical images (Othman et al. 2022), but also studies focusing on genetic data which employ random forest (Ye et al. 2003; Hasan et al. 2023). Lastly, in stomach cancer, popular convolutional and recurrent neural networks have been employed for automatic tumor detector from medical images. Kather et al. (2019); Zhao et al. (2022); Niu et al. (2020). These examples highlight the diversity of machine learning techniques that have been applied to different types of cancers, aiding in enhanced diagnosis, prognosis, prediction, and treatment planning.

It is interesting to realize that the 3 types of cancer with the highest impact and scientific production here analyzed in terms of ML research (breast, lung, and colon, in that order) are those with the highest number of cases worldwide in 2019: breast (19.8 million cases), colon and rectum (11.46 million cases), and lung (3.21 million cases) (Roser and Ritchie 2015). This fact indicates that ML-cancer research is applied proportionally to the most prevalent cancers at the present time. Interestingly, this correlation is not so clear between the ML scientific production and the number of deaths per year caused by each cancer type. Indeed, the most investigated cancer with ML tools (breast cancer) only produced a total of 700,660 deaths in 2019 (3.5% annual mortality), whereas colon and lung cancers showed higher mortality rates with 1.09 and 2.04 million deaths, respectively. Another important point with certain influence in the application of ML to cancer research is that ML methods usually serve as early detection and diagnosis techniques through image processing, and this task may perform more or less effectively for some specific cancer types than for others. For example, ML techniques are specially useful to predict lung or breast cancer detection with image processing (LG and AT 2013; Priya and Ramamurthy 2018). In any way, our results clearly show that the higher the incidence of a cancer type is, the greater the effort carried out in terms of ML studies on this type of cancer.

Regarding institutional affiliation (see Table 4), Harvard University leads in the number of publications for lung, colon, and breast cancer with 142, 95, and 177, respectively, being by far the one that published the most papers on the subject. The University of California System, Chinese Academy of Sciences, and the University of Texas System are of great importance as well since they are in the top 5 for up to 4 types of cancer. It is noteworthy that stomach cancer is mainly investigated by Asian institutions from China and Japan, monopolizing the top 5 (Chinese Academy of Sciences, National Cancer Center Japan, Shanghai Jiao Tong University, University Chinese Academy of Sciences CAS, and University of Tokyo). It is interesting to realize that, out of the 2.71 million total cases of stomach cancer in the world in 2019, Asia had an enormous amount of 2.09 million (almost 78% of the total cases) during that year (Roser and Ritchie 2015). In comparison, Europe, Africa, and America present 337,292; 57,830; and 225,947 cases, respectively. This high incidence explains the demonstrated leading role of Japanese and Chinese institutions in this type of cancer. In general terms, American, Chinese, and Japanese institutions account for the majority of research. Only Egypt's EKB breaks this rule by being the fifth institution with the most number of documents in breast cancer. This fact is surprising, since Egypt is not one of the countries with the highest number of cases.

Table 4 Scientific production on ML applied to different types of cancer classified by institution, country, and journal

Cancer Type	Institution	TSP	%TSP	Country	TSP	%TSP	Journal	TSP	%TSP
Lung	Harvard Univ	142	3.55	USA	1327	33.20	Medical Physics	127	3.18
	Univ. of California System	119	2.98	China	1117	27.95	Proceedings of Spie	93	2.33
	Univ. of Texas System	111	2.78	India	311	7.78	Scientific Reports	88	2.20
	Chinese Academy of Sciences	91	2.28	England	206	5.15	Frontiers in Oncology	85	2.18
	Harvard Medical School	83	2.08	South Korea	199	4.98	IEEE Access	73	1.83
Colon and rectum	Harvard Univ	95	4.45	USA	667	31.24	Proceedings of Spie	47	2.20
	Harvard Medical School	64	3.00	China	482	22.58	Scientific reports	45	2.11
	Univ. of California System	60	2.81	England	185	8.67	Cancers	36	1.69
	Univ. of Texas System	53	2.48	Japan	161	7.51	Lecture Notes in computer Science	36	1.69
	Massachusetts General Hospital	48	2.25	Germany	137	6.41	Medical Physics	32	1.50
Liver	Sun Yat Sen Univ	62	3.58	China	635	36.71	Frontiers in Oncology	45	2.60
	Chinese Academy of Sciences	58	3.35	USA	454	26.24	Medical Physics	34	1.97
	Fudan Univ	53	3.06	Japan	116	6.71	Scientific Reports	33	1.91
	Univ. of Texas System	51	2.95	India	106	6.13	Proceedings of Spie	29	1.68
	Zhejiang Univ	47	2.72	Germany	96	5.55	Cancers	27	1.56
Stomach	Chinese Academy of Sciences	28	3.80	China	297	40.35	Scientific Reports	20	2.72
	National Cancer Center Japan	21	2.86	USA	127	17.26	Gastrointestinal Endoscopy	18	2.45
	Shangai Jiao Tong Univ	20	2.72	Japan	110	14.95	World Journal of Gastroenterology	17	2.31
	Univ. Chin. Academy of Sci	20	2.72	South Korea	74	10.05	Digestive Endoscopy	14	1.90
	Univ. of Tokyo	20	2.72	Germany	36	4.89	Frontiers in Oncology	12	1.63
Breast	Harvard Univ	177	2.43	USA	2227	30.56	Proceedings of Spie	297	4.08
	Univ. of Texas System	170	2.33	China	1271	17.44	Lecture Notes in computer Science	197	2.70
	Univ. of California System	163	2.24	India	746	10.24	Scientific Reports	116	1.59
	Chinese Academy of Sciences	124	1.70	England	463	6.35	IEEE Access	103	1.41
	Egyptian Knowledge Bank EKB	117	1.61	Germany	307	4.21	Medical Physics	98	1.35

As expected from the above results, the USA and China have the highest number of published papers, accounting for the majority of the publications. 33.20, 31.24, 26.24, 17.26, and 30.56% of the papers published for lung, colon, liver, stomach, and breast cancer, respectively, belongs to the USA, while China contributes with 27.95, 22.58, 36.71, 40.35, and 17.44%, respectively, of the total number of published studies. The fact that the USA and China are two economic superpowers and invest the most in research and development can explain these results. But, in addition, they are the countries with the highest number of people with cancer in the world from 2017: 22.86 million cases (USA) and 22.42 million cases (China) (Roser and Ritchie 2015). England, India, Germany, and Japan also stand out, although their production is in general lower than the USA and China.

Journals in the field of computer science and medicine are the most commonly used to publish the studies analyzed here. Medical journals such as Medical Physics, Scientific Reports, Frontiers in Oncology, and Gastrointestinal Endoscopy are the most used journals for lung, colon, liver, and stomach cancer. Specifically, the journal Scientific Reports is the only one that appears among the top 5 in all the cancer types analyzed here. However, Proceedings of Spie leads studies for breast cancer. Although medical journals dominate for most of the cancers studied here, in total output, computer science journals have a leading role, thanks to the contribution of the big amount of documents published on breast cancer (Tables 2–3).

Knowing which countries and institutions publish the most in terms of ML applied to cancer is crucial for improving international collaboration between research groups specialized in the field. This study could substantially facilitate the search for expert researchers, groups, and institutions and thus the improvement, development, and advancement of research on this topic. In addition, knowing the journals where these papers are most published can facilitate the process of documentation, submission, and publication in the field.

Analysis of keywords co-occurrence on different types of cancers

Constructing network maps for co-occurrence keywords allowed us to visualize and evaluate the different global trends for machine learning studies on cancer. Five different network maps, one for each cancer type, were constructed and analyzed. For all the network maps, the minimum number of occurrences of a keyword was set at 30, in order to avoid spurious correlations. Also, the intensity in the correlation between the different keywords was expressed as TLS (total link strength) by Vos-viewer.

As expected, the most frequent co-occurrence keywords in all cancer types (see Fig. 3) were found to be important technical words in the fields, such as "cancer," "machine learning," "deep learning," and "survival." However, other frequent co-occurences reveal current trends in ML research applied to cancer. For example, it is worth mentioning "diagnosis" and other related terms such as "computer aided diagnosis," "detection," "computer aided detection," "prognosis," or "prediction." These results suggest that ML is mainly used as a diagnosis and prevention method and thus can be employed as an important tool for early detection of tumors, which plays a crucial step for disease treatment (Cruz and Wishart 2006). In addition, the high co-occurrence levels of the keyword "classification" indicated how the application of ML is also important in classifying the cancer disease by type, giving key information about how to proceed for the treatment. "Radiotherapy" and "chemotherapy" present as well a high co-occurrence in all cancer types analyzed here, suggesting that ML techniques are not only useful as an early diagnosis method, but also could play an important role in cancer treatment. Indeed, machine learning is being investigated to enhance the efficiency of radiotherapy treatments against tumors (Meyer et al. 2018; Deist et al. 2018). "Radiomics" is also highlighted as a keyword with high co-occurrence in all the analyzed cancer types. Radiomics is an artificial intelligence-based methodology which uses data-characterization algorithms to extract critical information from medical images through spatial distribution of signal intensities and pixel interrelationships. Many studies have revealed the potential of radiomics to improve clinical decision and radiotherapy workflow (Giraud et al. 2019; van Timmeren et al. 2020). However, still further development is necessary for the implementation of this technique in hospitals.

In the case of stomach or gastric cancer, we highlight the co-occurrence of the keyword *Helycobacter pylori* with a total amount of 43 (TLS=187) (see Fig. 3c). This highlights this bacterium as a major concern in stomach cancer. Indeed, a great number of studies about machine learning on stomach cancer are focused on people infected with the bacterium *H. pylori*, which could play a significant role in cancer development Polk et al. (2010). A similar case was found for liver cancer with the keyword *cirrhosis* (42 co-occurrences and TLS of 163) (Fig. 3d).

The keyword "benign" appeared most frequently for breast cancer with 118 total co-occurrences (TLS=749) in Fig. 3e. This suggests a higher appearance of benign tumors for this cancer type. This fact matches very well with the death rate for breast cancer, which has lower values (8.62 deaths per 100,000 individuals) in comparison with lung (25.18), colon (13.69), and stomach (11.88) cancer (Roser and Ritchie 2015). Interestingly, breast cancer is the only one that differentiates between patient's gender. The keyword woman reported a total of 233 co-occurrences with a TLS of 1142. However, male gender does not appear as an

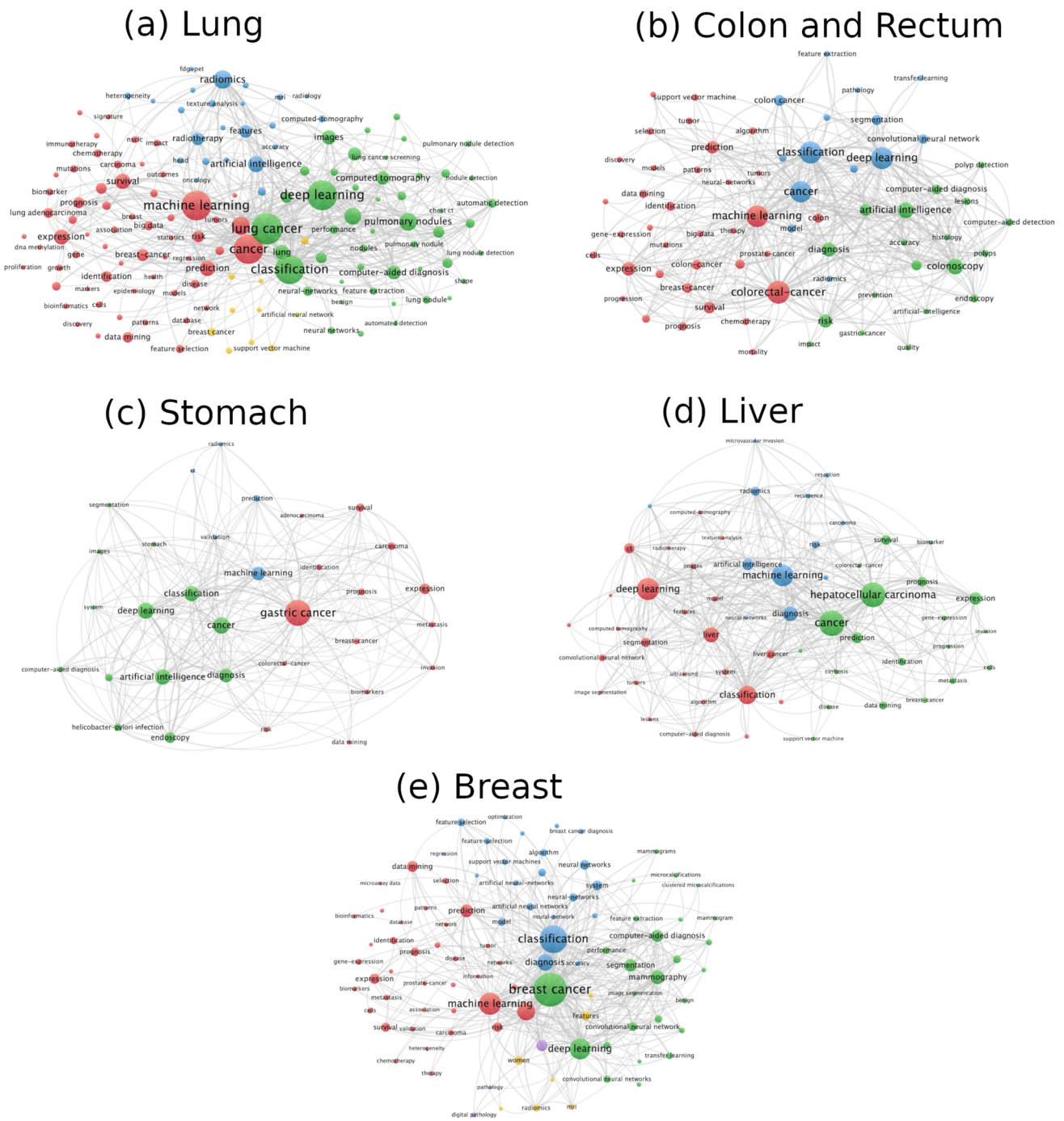

Fig. 3 Visualization network map of keywords co-occurrence for all the documents reported for machine learning on **a** lung, **b** colon and rectum, **c** stomach, **d** liver, and **e** breast cancer. The size of the spheres is proportional to the number of occurrences of each keyword. The lines represent the total link strength and the correlation between the keywords

important keyword on the cancer types analyzed within this work. This suggests most of ML studies on breast cancer are focused on female patients because most cases occur in women.

The co-occurence keyword analysis presented here clearly shows the current trends on machine learning studies in cancer such as the use of this technique in diagnosis, prognosis, detection, prediction, improvement of radiotherapy treatment, radiomics as a promising tool, etc. This analysis strongly supports the potential of these techniques and could be useful to boost further research in the field.

Machine learning in cancer research: a paradigm shift in diagnosis, treatment, and beyond

Along this section, we have analyzed a vast amount of literature discussing the potential of ML in cancer research. Here, we provide a brief summary of the main challenges and opportunities identified in this field, including ethical considerations. For further details, please refer to the provided references.

Early detection and accurate diagnosis. One of the most promising applications is improving early detection and accurate diagnosis. By training models on large-scale datasets comprising imaging data, clinical records, and genetic profiles, ML algorithms can aid in the identification of subtle cancer-related patterns that may escape human detection (McKinney et al. 2020).

Precision medicine and personalized treatment. Cancer is a highly heterogeneous disease, requiring tailored treatment approaches for each patient. ML enables the development of predictive models that consider individual patient characteristics, such as genetic variations and lifestyle factors, to guide personalized treatment decisions. By leveraging these models, clinicians can optimize therapy selection, dosage, and scheduling, leading to improved outcomes and reduced adverse effects (MacEachern and Forkert 2021).

Drug discovery and repurposing. Traditional drug discovery processes are time-consuming and costly. ML algorithms offer an innovative approach to accelerate the identification of potential therapeutic targets and drug candidates. By integrating large-scale molecular and clinical data, ML models can predict drug efficacy and toxicity, thereby enabling the identification of promising candidates for further experimental validation (Dara et al. 2022).

Challenges and ethical considerations. While ML holds immense promise, there are critical challenges that need to be addressed. Issues related to data quality, interpretability of complex models, bias in training datasets, and ethical considerations regarding patient privacy and consent require careful attention. Ensuring transparency, accountability, and equitable access to the benefits of ML-driven cancer research must be central to its implementation (Sorell et al. 2022; Yu et al. 2016).

In conclusion, ML has the potential to revolutionize cancer research. It represents a paradigm shift in our ability to harness the power of data and computational algorithms to confront the challenges posed by cancer. By addressing the associated challenges and ethical considerations, the integration of ML in cancer research holds immense promise for improving patient outcomes and transforming the landscape of cancer prevention and treatment.

Conclusions

In this work, we investigate the global scientific production and impact of machine learning applied to different cancer types. The huge amount of documents included in our study (more than 30,000), most of them published in the last few years, reveals the great interest raised recently in this field. The high levels of production in the publications that involve ML applications to cancer are a consequence of the combination of a novel and revolutionary technology such as ML and a decades-old research field such as cancer, with a deep impact on society and global health, causing millions of deaths annually.

By means of bibliometric methods, we have carried out a quantitative analysis of the scientific production on ML applied to the five most relevant cancer types (namely, breast, lung, colon/rectum, liver, and stomach cancer). We have confirmed that there exists a correlation between the incidence of different cancer types and the amount of publications that involve machine learning for that type of cancer. When classifying publications by countries, it is clear that the USA and China are the main scientific producers in the general case, but some local correlation can be found between the number of cases and number of publications, as in the case of China and Japan in the study of stomach cancer. Finally, the co-occurrence diagrams show intriguing correlations and point out the present and future trends of ML-cancer research, not only in the use of ML as an effective diagnostic method, but also as a useful tool for improving radiotherapy and chemotherapy-based treatments.

In addition to the interesting relations found when comparing the number of publications and the number of cancer cases, recognizing the countries and institutions that most study the field of ML applied to cancer can be helpful to establish international collaborations. Furthermore, knowledge of the journals where the studies are most published can

facilitate access to the appropriate information, as well as the process of submission and publishing in the field. Therefore, this work can serve as a guide for numerous researchers to get insights into the scientific production of ML applied to different cancer types, a remarkably active field due to its implications on global health.

Author Contributions MA Ruiz-Fresneda conceived the project and performed the bibliometric search. MA Ruiz-Fresneda, A Gijón, and P Morales-Álvarez analyzed the data, prepared figures and tables, and jointly wrote the paper.

Funding This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement no. 860627 (CLARIFY Project), from the Spanish Ministry of Science and Innovation under project PID2019-105142RB-C22, and by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades under the project P20_00286. Funding for open access charge: Universidad de Granada / CBUA.

Availability of data and materials Not applicable

Declarations

Consent for publication All authors agree that this research work can be published.

Consent to participate All authors declare that they have participated on this research work.

Ethical approval This declaration is not applicable for this work.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ain QU, Al-Sahaf H, Xue B, Zhang M (2022) Genetic programming for automatic skin cancer image classification. Expert Syst Appl 197:116680

Akay MF (2009) Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst Appl 36(2):3240–3247

Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. In 2017 international conference on engineering and technology (icet) (pp 1–6)

Aliferis CF, Hardin D, Massion PP (2002) Machine learning models for lung cancer classification using array comparative genomic hybridization. Proc AMIA Symp, 7–11

Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. Mol Pharm 13(7):2524–2530

American Cancer Society (2021) https://www.cancer.org/about-us/our-global-health-work/global-cancer-burden.html. (Accessed: June 2022)

Azari H, Nazari E, Mohit R, Asadnia A, Maftooh M, Nassiri M, Avan A (2023) Machine learning algorithms reveal potential miRNAs biomarkers in gastric cancer. Sci Rep 13(1):6147. https://doi.org/10.1038/s41598-023-32332-x

Bakrania A, Joshi N, Zhao X, Zheng G, Bhat M (2023) Artificial intelligence in liver cancers: decoding the impact of machine learning models in clinical diagnosis of primary liver cancers and liver cancer metastases. Pharmacol Res 189:106706

Cabral BP, da Graça Derengowski Fonseca M, Mota FB (2018) The recent landscape of cancer research worldwide: a bibliometric and network analysis. Oncotarget, 9. https://doi.org/10.18632/oncotarget.25730

Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A (2020) A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing 408:189–215

Chen H-L, Yang B, Liu J, Liu D-Y (2011) A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. Expert Syst Appl 38(7):9014–9022

Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, Tsirigos A (2018) Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat Med 24(10):1559-1567. Oct 01 Retrieved from https://doi.org/10.1038/s41591-018-0177-5

Cruz JA, Wishart DS (2006) Applications of machine learning in cancer prediction and prognosis. Cancer Informat 2. https://doi.org/10.1177/117693510600200030

Cuocolo R, Caruso M, Perillo T, Ugga L, Petretta M (2020) Machine learning in oncology: a clinical appraisal. Cancer Lett 481:55–62

Dara S, Dhamercherla S, Jadav SS, Babu CM, Ahsan MJ (2022) Machine learning in drug discovery: a review. Artif Intell Rev 55(3):1947–1999

Deepa P, Gunavathi C (2022) A systematic review on machine learning and deep learning techniques in cancer survival prediction. Prog Biophys Mol Biol 174:62-71. Retrieved from https://www.sciencedirect.com/science/article/pii/S0079610722000761 https://doi.org/10.1016/j.pbiomolbio.2022.07.004

Deist TM, Dankers FJ, Valdes G, Wijsman R, Hsu IC, Oberije C, Lambin P (2018) Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. Med Phys 45. https://doi.org/10.1002/mp.12967

Dhamija P, Bag S (2020) Role of artificial intelligence in operations environment: a review and bibliometric analysis. TQM Journal 32:869–896. https://doi.org/10.1108/TQM-10-2019-0243

dos Santos BS, Steiner MTA, Fenerich AT, Lima RHP (2019) Data mining and machine learning techniques applied to public health problems: a bibliometric analysis from 2009 to 2018. Comput Ind Eng 138. https://doi.org/10.1016/j.cie.2019.106120

Duan Y, Edwards JS, Xu M (2005) Web-based expert systems: benefits and challenges. Inf Manag 42(6):799–811

Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, Venâncio R (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 318(22):2199–2210. https://doi.org/10.1001/jama.2017.14585

Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR (2018) Deep learning for healthcare applications based on physiological signals: a review. Comput Methods Prog Biomed 161:1–13. https://doi.org/10.1016/j.cmpb.2018.04.005

Fiste O, Liontos M, Zagouri F, Stamatakos G, Dimopoulos MA (2022) Machine learning applications in gynecological cancer: a critical review. Crit Rev Oncol Hematol 179:103808. Retrieved from https://www.sciencedirect.com/science/article/pii/S1040842822002323 https://doi.org/10.1016/j.critrevonc.2022.103808

Gayathri BM, Sumathi CP (2016) Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer. In 2016 IEEE international conference on computational intelligence and computing research (iccic) (p 1-5) https://doi.org/10.1109/ICCIC.2016.7919576

Giraud P, Giraud P, Gasnier A, El Ayachy R, Kreps S, Foy J-P, Bibault J-E (2019) Radiomics and machine learning for radiotherapy in head and neck cancers. Frontiers in Oncology, 9. Retrieved from https://www.frontiersin.org/article/10.3389/fonc.2019.00174 https://doi.org/10.3389/fonc.2019.00174

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press

Hameed N, Shabut AM, Ghosh MK, Hossain MA (2020) Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. Expert Syst Appl 141:112961

Hasan ME, Mostafa F, Hossain MS, Loftin J (2023) Machine-learning classification models to predict liver cancer with explainable AI to discover associated genes. AppliedMath 3(2):417–445. Retrieved from https://www.mdpi.com/2673-9909/3/2/22 https://doi.org/10.3390/appliedmath3020022

Houssein EH, Emam MM, Ali AA, Suganthan PN (2021) Deep and machine learning techniques for medical imaging-based breast cancer: a comprehensive review. Expert Syst Appl 167:114161

Isaksson LJ, Pepa M, Zaffaroni M, Marvaso G, Alterio D, Volpe S, Jereczek-Fossa BA (2020) Machine learning-based models for prediction of toxicity outcomes in radiotherapy. Frontiers in Oncology, 10. Retrieved from https://www.frontiersin.org/article/10.3389/fonc.2020.00790 https://doi.org/10.3389/fonc.2020.00790

Jamwal A, Agrawal R, Sharma M, Kumar A, Kumar V, Garza-Reyes JAA (2021) Machine learning applications for sustainable manufacturing: a bibliometric-based review for future research. J Enterp Inf Manag. https://doi.org/10.1108/JEIM-09-2020-0361

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. CA Cancer J Clin 61. https://doi.org/10.3322/caac.20107

Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349. https://doi.org/10.1126/science.aaa8415

Joshi SA, Bongale AM, Bongale A (2021) Breast cancer detection from histopathology images using machine learning techniques: a bibliometric analysis. Libr Philos Pract, 2021

Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, Luedde T (2019) Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 25(7):1054-1056. Jul 01 Retrieved from https://doi.org/10.1038/s41591-019-0462-y

Kennion O, Maitland S, Brady R (2022) Machine learning as a new horizon for colorectal cancer risk prediction? A systematic review. Health Sci Rev 4:100041. Retrieved from https://www.sciencedirect.com/science/article/pii/S2772632022000307 https://doi.org/10.1016/j.hsr.2022.100041

Koppad S, Basava A, Nash K, Gkoutos GV, Acharjee A (2022) Machine learning-based identification of colon cancer candidate diagnostics genes. Biology 11(3). Retrieved from https://www.mdpi.com/2079-7737/11/3/365 https://doi.org/10.3390/biology11030365

Lachman R (1989) Expert systems: a cognitive science perspective. Behav Res Methods Instrum Comput 21(2):195–204

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444

Lg A, At E (2013) Using three machine learning techniques for predicting breast cancer recurrence. J Health Med Inform 04. https://doi.org/10.4172/2157-7420.1000124

Li Y, Wu X, Yang P, Jiang G, Luo Y (2022) Machine learning for lung cancer diagnosis, treatment, and prognosis. Genomics Proteomics Bioinformatics 20(5):850–866

López-Pérez M, Amgad M, Morales-Álvarez P, Ruiz P, Cooper LA, Molina R, Katsaggelos AK (2021) Learning from crowds in digital pathology using scalable variational gaussian processes. Scientific Reports 11(1):1–9

López-Úbeda P, Díaz-Galiano MC, Martín-Noguerol T, Ureña-López A, Martín-Valdivia M-T, Luna A (2020) Detection of unexpected findings in radiology reports: a comparative study of machine learning approaches. Expert Syst Appl 160:113647

MacEachern SJ, Forkert ND (2021) Machine learning for precision medicine. Genome 64(4):416–425

Majumder SK, Ghosh N, Gupta PK (2005) Relevance vector machine for optical diagnosis of cancer. Lasers Surg Med 36(4):323–333. https://doi.org/10.1002/lsm.20160

Maurya S, Tiwari S, Mothukuri MC, Tangeda CM, Nandigam RNS, Addagiri DC (2023) A review on recent developments in cancer detection using machine learning and deep learning models. Biomedical Signal Processing and Control 80:104398. Retrieved from https://www.sciencedirect.com/science/article/pii/S1746809422008527 https://doi.org/10.1016/j.bspc.2022.104398

McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H et al (2020) International evaluation of an AI system for breast cancer screening. Nature 577(7788):89–94

Meena DK, Mathaiyan J (2021) Essential medicines research in India: situation analysis. Journal of Young Pharmacists 13(2):82

Meyer P, Noblet V, Mazzara C, Lallement A (2018) Survey on deep learning for radiotherapy. Comput Biol Med 98. https://doi.org/10.1016/j.compbiomed.2018.05.018

Mirmozaffari M (2019) Presenting a medical expert system for diagnosis and treatment of nephrolithiasis. European Journal of Medical and Health Sciences 1. https://doi.org/10.24018/ejmed.2019.1.1.20

Mokoatle M, Marivate V, Mapiye D, Bornman R, Hayes V et al (2023) A review and comparative study of cancer detection using machine learning: SBERT and SimCSE application. BMC bioinformatics 24(1):1–25

Morales-Álvarez P, Ruiz P, Coughlin S, Molina R, Katsaggelos AK (2022) Scalable variational Gaussian processes for crowdsourcing: glitch detection in LIGO. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(3):1534–1551. https://doi.org/10.1109/TPAMI.2020.3025390

Muhuri PK, Shukla AK, Abraham A (2019) Industry 4.0: a bibliometric analysis and detailed overview. Eng Appl Artif Intell 78. https://doi.org/10.1016/j.engappai.2018.11.007

Munim ZH, Dushenko M, Jimenez VJ, Shakil MH, Imset M (2020) Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. Maritime Policy and Management 577–597. https://doi.org/10.1080/03088839.2020.1788731

Murphy KP (2012). Machine learning: a probabilistic perspective. MIT press

Nathan R, Monk CT, Arlinghaus R, Adam T, Alós J, Assaf M et al (2022) Big-data approaches lead to an increased understanding of the ecology of animal movement. Science 375(6582):eabg1780

Nguyen T-L, Kavuri S, Park S-Y, Lee M (2022) Attentive hierarchical ANFIS with interpretability for cancer diagnostic. Expert Syst Appl 201:117099

Niu P-H, Zhao L-L, Wu H-L, Zhao D-B, Chen Y-T (2020) Artificial intelligence in gastric cancer: application and future perspectives. World J Gastroenterol 26(36):5408–5419

Othman E, Mahmoud M, Dhahri H, Abdulkader H, Mahmood A, Ibrahim M (2022) Automatic detection of liver cancer using hybrid pre-trained models. Sensors (Basel) 22(14)

Paluszynski W, Kalet I, Laramore G, Borning A (1983) An artificial intelligence program to plan radiotherapy for cancer of the oral cavity. In Medical physics (Vol 10. pp 739–739)

Peng T, Gu Y, Ye Z, Cheng X, Wang J (2022) A-LugSeg: automatic and explainability-guided multi-site lung detection in chest X-ray images. Expert Syst Appl 198:116873

Polk DB, Peek RM (2010) Helicobacter pylori: gastric cancer and beyond. Nat Rev Cancer 10. https://doi.org/10.1038/nrc2857

PR, R, Nair RA, G V (2019) A comparative study of lung cancer detection using machine learning algorithms. In 2019 IEEE international conference on electrical, computer and communication technologies (icecct) (p 1-4). https://doi.org/10.1109/ICECCT.2019.8869001

Priya SS, Ramamurthy B (2018) Lung cancer detection using image processing techniques. Research Journal of Pharmacy and Technology 11. https://doi.org/10.5958/0974-360X.2018.00379.7

Ravuri M, Kannan A, Tso GJ, Amatriain X (2018). Learning from the experts: from expert systems to machine-learned diagnosis models. In Machine learning for healthcare conference (pp 227–243)

Roser M, Ritchie H (2015) Our world in data - cancer. https://ourworldindata.org/cancer. (Accessed: June 2022)

Saibene A, Assale M, Giltri M (2021) Expert systems: definitions, advantages and issues in medical field applications. Expert Syst Appl 177:114900

Salod Z, Singh Y (2020) A five-year (2015 to 2019) analysis of studies focused on breast cancer prediction using machine learning: a systematic review and bibliometric analysis. J Public Health Res 9. https://doi.org/10.4081/jphr.2020.1772

Shilo S, Rossman H, Segal E (2020) Axes of a revolution: challenges and promises of big data in healthcare. Nat Med 26(1):29–38

Shortliffe EH (1986) Medical expert systems–knowledge tools for physicians. West J Med 145(6):830

Silva-Rodríguez J, Colomer A, Sales MA, Molina R, Naranjo V (2020) Going deeper through the Gleason scoring scale: an automatic end-to-end system for histology prostate grading and cribriform pattern detection. Comput Methods Prog Biomed 195:105637

Sirinukunwattana K, Raza SEA, Tsang Y-W, Snead DRJ, Cree IA, Rajpoot NM (2016) Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans Med Imaging 35(5):1196–1206. https://doi.org/10.1109/TMI.2016.2525803

Sorell T, Rajpoot N, Verrill C (2022) Ethical issues in computational pathology. J Med Ethics 48(4):278–284

Stout NL, Alfano CM, Belter CW, Nitkin R, Cernich A, Siegel KL, Chan L (2018) A bibliometric analysis of the landscape of cancer rehabilitation research (1992–2016). J Natl Cancer Inst 110. https://doi.org/10.1093/jnci/djy108

Tabl AA, Alkhateeb A, ElMaraghy W, Rueda L, Ngom A (2019) A machine learning approach for identifying gene biomarkers guiding the treatment of breast cancer. Front Genet 10. Retrieved from https://www.frontiersin.org/article/10.3389/fgene.2019.00256 https://doi.org/10.3389/fgene.2019.00256

Tran BX, McIntyre RS, Latkin CA, Phan HT, Vu GT, Nguyen HLT, Ho RC (2019) The current research landscape on the artificial intelligence application in the management of depressive disorders: a bibliometric analysis. International Journal of Environmental Research and Public Health 16. https://doi.org/10.3390/ijerph16122150

Tschuggnall M, Grote V, Pirchl M, Holzner B, Rumpold G, Fischer MJ (2021) Machine learning approaches to predict rehabilitation success based on clinical and patient-reported outcome measures. Informatics in Medicine Unlocked 24:100598

Turki T (2018) An empirical study of machine learning algorithms for cancer identification. 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 1-5

Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G et al (2019) Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 18(6):463–477

van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B (2020) Radiomics in medical imaging—"how-to" guide and critical reflection. Insights into Imaging 11:91. Retrieved from https://doi.org/10.1186/s13244-020-00887-2

Vokinger KN, Gasser U (2021) Regulating AI in medicine in the United States and Europe. Nature machine intelligence 3(9):738–739

Wang K, Feng C, Li M, Pei Q, Li Y, Zhu H, Tan F (2020) A bibliometric analysis of 23,492 publications on rectal cancer by machine learning: basic medical research is needed. Ther Adv Gastroenterol 13. https://doi.org/10.1177/1756284820934594

Wang K, Pleiss G, Gardner J, Tyree S, Weinberger KQ, Wilson AG (2019) Exact gaussian processes on a million data points. Advances in Neural Information Processing Systems, 32

Wang L (2022) Deep learning techniques to diagnose lung cancer. Cancers 14(22). Retrieved from https://www.mdpi.com/2072-6694/14/22/5569 https://doi.org/10.3390/cancers14225569

WHO (2022) World Health Organization. https://www.who.int/news-room/fact-sheets/detail/cancer. (Accessed: June 2022)

WOS (2022) Web of Science - WOS Core Collection. https://www.webofscience.com/wos/woscc/basic-search. (Accessed: June 2022)

Xu Z, Yu D, Wang X (2019) A bibliometric overview of international journal of machine learning and cybernetics between 2010 and 2017. International Journal of Machine Learning and Cybernetics 10:2375–2387. https://doi.org/10.1007/s13042-018-0875-9

Ye Q-H, Qin L-X, Forgues M, He P, Kim JW, Peng AC, Wang XW (2003) Predicting hepatitis b virus–positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. Nat Med 9(4):416-423. Apr 01 Retrieved from https://doi.org/10.1038/nm843

Yu C, Helwig EJ (2021) The role of AI technology in prediction, diagnosis and treatment of colorectal cancer. Artif Intell Rev 55(1):323–343

Yu K-H, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, Snyder M (2016) Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 7(1):12474

Yu Y, Si X, Hu C, Zhang J (2019) A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput 31(7):1235–1270

Zhang Y, Chen J-H, Lin Y, Chan S, Zhou J, Chow D, Su M-Y (2020) Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers. Eur Radiol 31(4):2559–2567

Zhao Y, Hu B, Wang Y, Yin X, Jiang Y, Zhu X (2022) Identification of gastric cancer with convolutional neural networks: a systematic review. Multimedia Tools and Applications 81(8):11717–11736

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.