International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 12 | Dec 2024 www.irjet.net p-ISSN: 2395-0072

# A Bibliometric Analysis of Artificial Intelligence for Cancer Detection and Research Directions

**RAKSHITHA C M[1] ,RACHANA B G[2] ,SMRUTHI M S[3] ,VINDHYA V ALUR[4]**

[1 2 3 4] *Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and Technology Davangere, affiliated to VTU Belagavi, Karnataka, India*

---

**Abstract -** Cancer is the leading cause of mortality globally, second only to cardiovascular disorders. Early cancer disease detection greatly increases the likelihood of recovery. Artificial intelligence is one set of technology that is being used more and more to identify cancer. Because it enables the early detection of carcinomas, artificial intelligence has significant promise for helping physicians and other healthcare professionals. Artificial intelligence research for cancer diagnosis has expanded significantly in recent years. We carried out a bibliometric analysis of the body of literature pertaining to the use of AI in cancer diagnosis for this article. 6450 articles published between 1986 and 2022 on that subject were examined by us. We were able to provide an overview of this research field by doing this, covering its main subjects, pertinent publications, organizations, and articles. We created a study agenda for the future based on our findings, which may contribute to the advancement of artificial intelligence research for cancer diagnosis. In conclusion, our research aims to provide a platform and basis for scientists who are curious in the possibilities of artificial intelligence in cancer detection.

***Key Words*: cancer detection; artificial intelligence; machine learning; deep learning; bibliometric study.**

## 1.INTRODUCTION

The fundamental building blocks of all plants and animals are living cells. These cells proliferate continuously to allow for growth or to replace damaged cells. Even though this process is typically balanced and under control, disruption to this genetic control may result in cancer [1]. Most cell-based life is susceptible to the condition known as cancer. It affects humanity for as long as it has existed, and the ancient Egyptians were the first to notice it [2]. Globally, cancer is the leading cause of death, second only to cardiovascular illnesses [3]. Globally, there were 9.6 million cancer-related deaths and over 18 million new cases of cancer in 2018 [4]. Considering the danger that cancer poses, scientists have long attempted to determine ways to To improve the odds of recovery, it's critical to identify cancer as soon as possible in addition to receiving therapy once it manifests [5–8]. Lung cancer is the most deadly type of cancer in part because it is hard to find in its early stages and difficult to treat when it has progressed [9, 10]. New strategies are always being developed to support an early cancer diagnosis because of the significant advantages of early cancer detection.Since its introduction in 1960 [11], mammography has become one of the most widely used methods for detecting breast cancer [12]. Computers are being utilized more and more to assist clinical practitioners in making medical diagnoses as a result of digitization and increases in processing capacity. Computer programs that aid in cancer detection Artificial intelligence (AI) is one technology that has gained more attention recently. The term artificial intelligence (AI) encompasses a wide range of innovations and technology. AI has been used in medicine recently for a number of reasons, such as assisting medical professionals in making decisions [15]. AI is being studied and applied more and more in the field of oncology for a variety of reasons [14]. The identification and diagnosis of cancer is one potential use. AI has the potential to revolutionize the early identification of cancer disorders and is the next development in CAD because of its ability to efficiently screen for or diagnose cancer or polyps [14,16].

AI for cancer detection and diagnosis has expanded quickly in recent years, both in clinical practice and as a field of study. Research on AI-supported cancer diagnosis has been growing yearly since the 2010s. These days, a variety of disciplines, including engineering, computer science, medicine, and mathematics, contribute to this research area. A thorough study that attempts to provide an overview of the research field of AI in cancer detection as a whole does not exist, despite the fact that there are several studies regarding AI on cancer [17–19]. This is surprising because AI for cancer diagnosis is now a vast and challenging topic to monitor because to the abundance of research and publications. This makes it challenging for scholars who are interested.

## 2. Artificial Intelligence Foundations

The first artificial neuron notion was put forth by [28] in 1943 [27], marking the beginning of artificial intelligence. The phrase "artificial intelligence" was first used at the Darthmouth Conference thirteen years later [29]. As a result, artificial

intelligence (AI) is one of the newest areas being studied in science and engineering [29]. It is currently a vibrant and complicated discipline with a wide range of study topics as well as various use cases and applications for businesses and in practice [30–32]. AI has grown significantly in recent years and is being considered with attention by both society and industry. The primary causes are increases in processing power and the amount of data accessible for AI system training [33]. It is crucial to remember that artificial intelligence is a multidisciplinary field that is being studied in a number of academic disciplines, such as computer science, psychology, neurology, and mathematics [34, 35]. The term artificial intelligence (AI) encompasses a wide range of algorithms and technologies.

Artificial neural networks are among the most widely utilized AI technology. Deep learning is another term for artificial neural networks that are multilayered and comprise multiple hidden layers [30, 31]. The goal of artificial neural networks is to mimic how biological beings, including humans, learn [36]. Because of this, artificial neural networks are modeled after the brains of biological things and are made up of interconnected processing units called neurons [31]. After receiving inputs, these neurons process them in accordance with predetermined guidelines to produce an output. These neurons are frequently organized into many modules or layers. The phrase "deep learning" in this sense refers to AI encompasses a number of additional technologies, including support vector machines [39,40] and random forests [37,38]. However, an explanation of these technologies would be outside the purview of this paper and is not required to comprehend the study's subsequent findings. Modern AI systems are not intelligent in the strict sense, despite their many capabilities. The distinction between strong and weak AI was first made by [41] in order to characterize the capabilities of AI. Weak artificial intelligence (AI) systems are typically not intelligent and are solely designed for specific tasks. They also lack other aspects of humanity, such as sentiments, emotions, and a conscious mind [34, 41]. Even though weak AI systems frequently give the impression of being clever, they simply act that way [29, 42]. On the other hand, AI systems that possess human-like intelligence or capacities are referred to as strong AI, sometimes known as artificial general intelligence (AGI) [43, 44]. In addition to intelligence, this may also imply that these systems are emotional or sentimental [34].

## 3. Approach

This section explains our bibliometric methodology. A bibliometric study can be conducted in two steps, broadly speaking. The first step is to gather the data that will be evaluated. The first subsection explains this step. The actual data analysis phase comes after the data collection phase. The second subsection provides an outline of this procedure.

### 3.1. Data Collection

To begin our investigation, the bibliometric data had to be gathered. Nowadays, there are a number of databases available for gathering bibliometric data, with Web of Science and Scopus being two of the most often used [47, 48]. The functions and characteristics of these databases vary [49]. We choose to adhere to [20]'s advice and exclusively gather bibliometric data from a single database. We decided to gather our data from the scientific database Scopus. The well-known database Scopus has been employed in a number of previous bibliometric research [21,47,48,5052]. Furthermore, Scopus has a larger journal coverage than Web of Science, making it a good tool for finding as much research as possible [26]. We chose not to use other databases, such as PubMed and Google Scholar, despite their existence. In contrast to Google Scholar, Scopus offers the ability to immediately download all bibliometric metadata and create a comprehensive search string. Second, Scopus includes far more interdisciplinary research than PubMed. Because AI-based cancer detection is a diverse research field, we determined that the best database to perform a bibliometric analysis on was Scopus.

The technical section includes broad technical words such as "machine learning" and "artificial intelligence." In an effort to expand our search, we also looked for particular technologies, including "artificial neural network," "deep learn*," "fuzzy expert system," or "evolutionary computation." The phrases used for the application were "cancer diagnose*" and "cancer detect*." The * symbol is used because of Scopus' syntax, which enables searching for every potential word ending for the search phrase.

### 3.2. Analysis of Data

Numerous tools that aid in the analysis of bibliometric data have emerged in recent years [20].We combined two tools—Bibliometrix/Bilioshiny and VOSviewer—in our investigation. Bibliometrix is an open source R package that was created by [54]. It makes it possible to analyze bibliometric data in a wide range of ways [49]. Additionally, we added Biblioshiny to Bibliometrix. We have added VOSviewer to complement Biblioshiny and Bibliometrix. One tool for visualizing bibliometric data is VOSviewer.It was created by the Centre for Science and Technology Studies at Leiden University in the Netherlands [49,55]. Numerous bibliometric studies have used VOSviewer, which makes it possible to create bibliometric networks that display connections between publications, outlets, keywords, and researchers, among other things.

Furthermore, co-authorship analysis, bibliographic coupling, and co-citation creation are supported by VOSviewer [49,55].VOSviewer is a good tool for visualizing keyword co-occurrences, even if Biblioshiny has more statistical features.

## 4.Overview and General Metrics

We will give a summary of our sample and some broad data, like yearly production, document kinds, and contributing author details, in this first subsection. An overview of the fundamental metrics of our final sample is provided in Table 1. The sample comprises 6450 distinct papers in total. 23,854 distinct scholars have written and co-written these documents, meaning that each author has contributed 0.270 documents.9321 author keywords and 247,762 references were cited overall. More than 21,192 keywords were also found. The 6450 papers, which came from various sources and were published in 2018, averaged 19.87 citations. About 25% of the 6232 articles with multiple authors were created by a global team. The topic of this study is timely because of the The average document age is only 3.72 years old, which supports the topic's timeliness. This suggests that most of the research has been released during the last four years.

**Table 1**: Key data and broad measurements.

| METRIC | VALUES |
|---|---|
| **Main information** | |
| Sources (conferences and journals) | 2018 |
| Documents | 6450 |
| Timespan of publications | 1986–2022 |
| Average citations per document | 19.87 |
| Average document age | 3.72 |
| Total number of references | 247,762 |
| Number of author's keywords | 9321 |
| Number of keywords plus | 21,192 |
| Document types | |
| Journal article | 4016 |
| Conference article | 1729 |
| Review | 708 |
| Authors and collaboration | |
| Number of different AI-cancer authors | 23,854 |
| Documents per AI-cancer author | 0.270 |
| Single-authored documents | 218 |
| Multi-authored documents | 6232 |
| Authors of multi-authored documents | 23,651 |
| Co-authors per document | 5.89 |
| Collaboration index | 3.8 |
| International co-authorship | 24.97% |

We contrasted our bibliometric information with that of existing bibliometric research on several subjects (see to Table 2 for a summary). First, it is notable that only a small percentage of articles on AI for cancer diagnosis have been written by a single person. Of the 6450 documents, only 218 (3.38%) were single-authored articles. This might be an evidence of the very high intricacy of this topic that makes it necessary to work together in huge author teams. The high collaboration index

for our investigation supports this idea even further. The collaboration index, which is computed by dividing the total number of authors who contributed to multi-authored papers by the total number of multi-authored articles, is frequently used to gauge researcher cooperation [56,57]. When compared to other bibliometric studies, this one has the fewest documents per author. This demonstrates that the topic of AI for cancer diagnosis is not dominated by a small number of researchers and that many different researchers contribute to it.

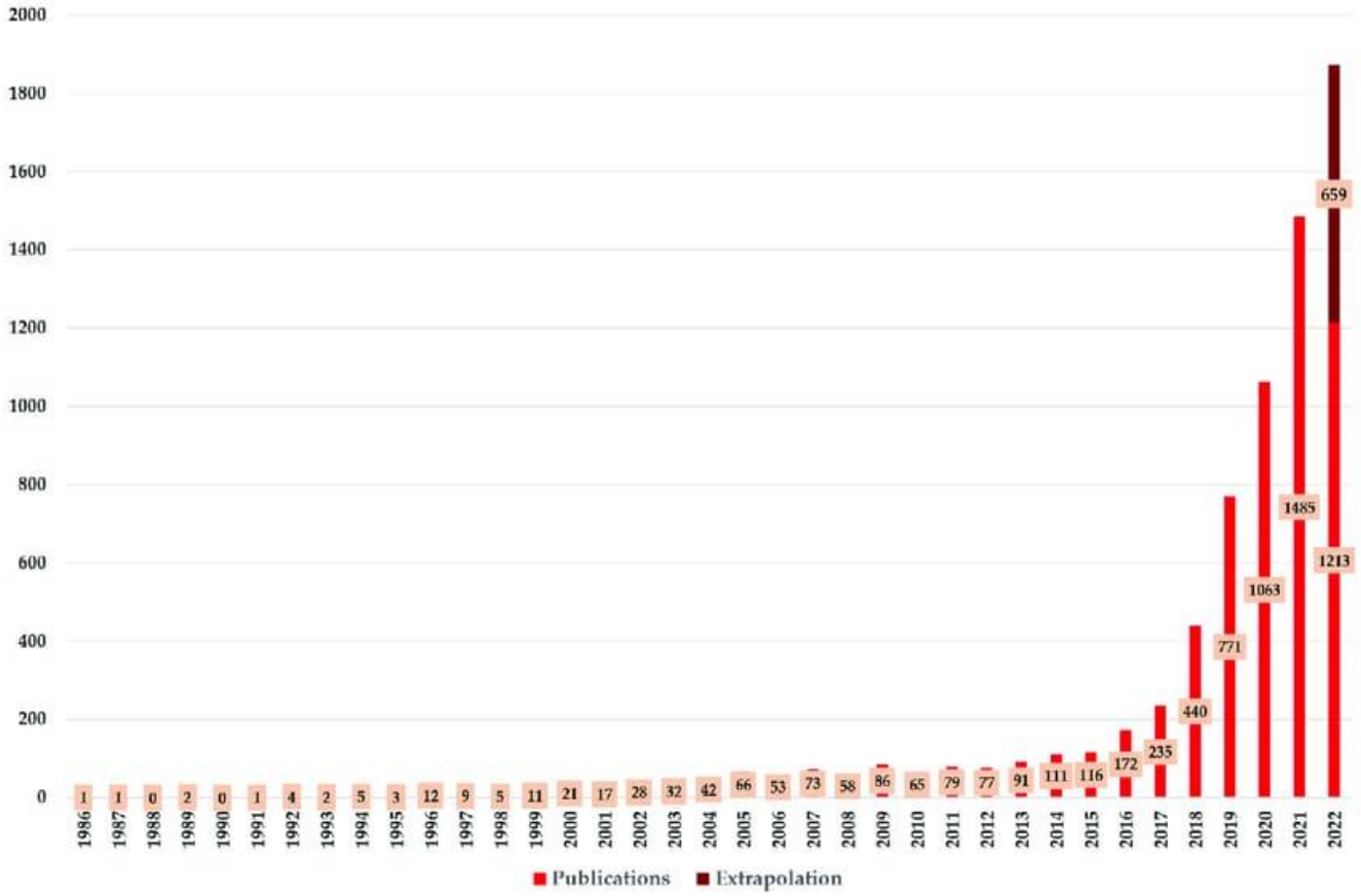


**Fig-1**. Overview of the annual production.

Fig-1 shows the annual production of research dealing with AI for cancer detection. The 1980s saw the publication of the first study on the subject. The initial article was published in 1986. An expert system for cervical cancer early detection was put forth in this article [60]. The yearly production of AI for cancer diagnosis increased very little until 1995. There were no publications on this subject in 1988 or 1990. The number of publications only increased modestly in the ensuing years. In 2014, the annual productions reached the 100-mark for the first time with 111 publications. AI has become more relevant in the realm of cancer diagnosis as its significance and possibilities have grown overall. Consequently, the majority of publications have been released in the last few years (2019–2021).Prior to the data collection for our study on September 23, 2022, 1213 publications had already been published. We anticipate that the trend of rising publications will continue in 2022 since it seems statistically that more articles are released in the final months of a year [21]. An extrapolation leads us to believe that there will be 1872 publications in total in 2022, with an estimated 659 pieces released after September 23.

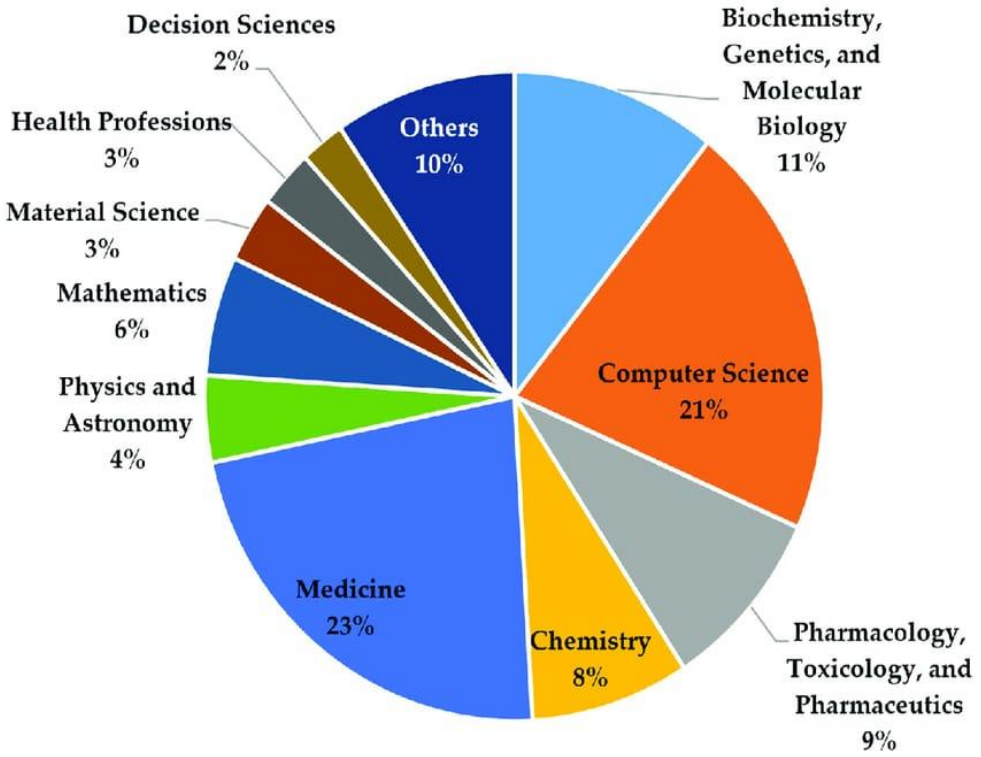


**Fig-2**. Overview of the most contributing disciplines.

The distribution of disciplines throughout the articles is seen in Figure 3. Figure 3's data came from Scopus, which assigns publications to disciplines according to the journal in which they were published. Nonetheless, some conferences or periodicals may fall under more than one field. Not surprisingly, the most widely used AI applications for cancer diagnosis are in the fields of medicine and computer science. These disciplines' venues have published 23% and 21% of all articles, respectively. Chemistry (8%), pharmacology, toxicology, and pharmaceutics (9%), biochemistry, genetics, and molecular biology (11%), and medicine and computer science come next. It is not unexpected that computer science and medicine dominate, as oncology and the identification and management .

## 5. Future Research Agenda

In the preceding sections, we presented the results of our bibliometric investigation. Based on our results, we will suggest intriguing avenues for future study in this section. These are intended to serve as a guide for interested scholars. As can be seen by first examining the word cloud in Figure 5 and the most-cited papers, present research is mostly focused on the predictive performance of a limited number of applied AI algorithms. The topic of human–computer interaction, or the interaction between the computer system and its users, receives a lot less attention. It is essential to consider how AI and human interaction may or should seem in the context of cancer diagnosis. In general, there are three potential outcomes: augmentation, substitution, and assemblage [95,96].

Future research must examine how AI and humans may work together most effectively in the context of cancer diagnosis. This means figuring out whether a switch is desirable at all and, more crucially, whether it is possible. There are already several promising studies on human–computer interaction in the healthcare industry [97,98]. These results can therefore be used as a foundation for further investigation into the relationship between humans and AI. The degree of trust between the public and the AI cancer detection system is another important factor.

**Table 2.** Future research agenda

| Focus | Possible Research Questions |
|---|---|
| Human Computer Interaction | How can the interaction between doctors and AI models be designed efficiently? What is the current state of trust towards AI based models in medicine?<br>How can trust in AI be built for doctors and patients?<br>How can AI experts and clinical practitioners cooperate and work together in the best way?<br>What is the role of explainable AI for building trust? |
| Robustness and security | How reliable are trained AI models on other cancer datasets (e.g., generated by other sensors)?<br>Can adversarial attacks outsmart AI models in medicine?<br>How should an AI system for cancer detection be designed to make it robust and secure against adversarial attacks and actors?<br>Could a cancer detection algorithm be applied to other types of cancer? |
| Explainable AI | Should explainable AI models be preferred instead of the most accurate one?<br>How should explainable AI be designed to increase the trust in the AI system and its decisions?<br>What are the promising approaches in XAI that have not yet been applied in the medical field? |
| Data Storage | Where should the data of the scans be stored to ensure data privacy rights?<br>Do new technologies like blockchain have a potential for the storage and management of medical data?<br>Should patient data be irreversibly anonymized or only pseudonymized? |

## 6. Conversation

We performed a bibliometric analysis of 6450 publications in this study that addressed the potential and use of AI in cancer diagnosis and detection. To the best of our knowledge, this is the first study to examine AI research for cancer diagnosis using a bibliometric approach. Both scholars and practitioners can profit from this work in a number of ways. The paper at hand can first be used by interested researchers to get a preliminary summary of the research on AI for cancer diagnosis. This includes details about the state of science, the most important papers, and the main subjects and issues that have been studied. Therefore, this study can help interested scholars gain a basic understanding of the research field pertaining to AI-based cancer detection; our research agenda can also serve as a basis for future research to build upon in order to further develop this exciting field; and clinical and commercial practitioners can use our study to gain a preliminary understanding of the potential of AI for supporting cancer diagnosis. However, we believe that most of our key findings, especially the most important subjects and themes, would likely not alter even if additional databases were used. Moreover, other bibliometric tools and analytic methods, like bibliographic coupling [122,123] or citation analysis [121], may be used to get additional data not covered in this work. The field of AI is likewise one that is expanding quickly.New research on AI-assisted cancer diagnosis is published each month. Consequently, the results of this investigation can only illustrate the current state of the art.

## 7. Results

The findings of our bibliometric analysis are presented in the next three parts. We will start by providing a broad summary of the sample we gathered and displaying the essential critical metrics. The outcomes of our performance analysis will then be presented. An overview of the sources with the most papers on AI for cancer detection may be found in the first section.Second, we list the nations, funding sponsors, and affiliations that contribute the most.We provide a thematic analysis of the most pertinent subjects and major themes following the performance analysis.

## 8. CONCLUSIONS

Artificial intelligence is a promising technology that is increasingly being utilized to detect or diagnose cancer.Numerous studies have been conducted on the topic as a result of the rapid expansion of research on AI for cancer diagnosis in recent years. It could be difficult for interested academics or medical professionals to acquire a basic comprehension of this topic because there is so much research available. Given this, our objective was to provide researchers with an overview and analysis of the field of study on AI for cancer diagnosis. To accomplish this, we conducted a bibliometric analysis of the literature on the topic. . Our analysis consisted of several parts. We started by giving a general overview of our sample and describing how scientific output changed over the course of the year and the disciplines that contributed to it. After that, we conducted a performance analysis. The countries and organizations with the highest production were listed here. We found that the two cancer kinds that recent studies have treated the most Based on these results, we developed a research agenda for the future that should aid scientists in advancing the field of AI-based cancer diagnosis.

## REFERENCES

[1] Kessler,M.M.Bibliographiccouplingbetweenscientificpapers.Am.Doc.1963,14,10–25.[CrossRef]

[2] Cruz, J.A.; Wishart, D.S. Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Inform. 2006, 2, 59–77.[CrossRef]

[3] Lu, G.; Fei, B. Medical hyperspectral imaging: A review. J. Biomed. Opt. 2014, 19, 010901. [CrossRef]

[4] Donthu, N.; Kumar, S.; Pattnaik, D. Forty-five years of Journal of Business Research: A bibliometric analysis. J. Bus. Res. 2019, 109, 1–14. [CrossRef]

[5] Moral-Muñoz, J.A.; Herrera-Viedma, E.; Santisteban-Espejo, A.; Cobo, M.J. Software tools for conducting bibliometric analysis in science: An up-to-date review. Prof. Inf. 2020, 29, e290103. [CrossRef]

[6] Adams, S.; Arel, I.; Bach, J.; Coop, R.; Furlan, R.; Goertzel, B.; Hall, J.S.; Samsonovich, A.; Scheutz, M.; Schlesinger, M.; et al.Mapping the Landscape of Human-Level Artificial General Intelligence. AI Mag. 2012, 33, 25–42. [CrossRef]

[7] Franklin, S. History, motivations, and core themes. In The Cambridge Handbook of Artificial Intelligence; Frankish, K., Ramsey, W.M.,Eds.; Cambridge University Press: Cambridge, UK, 2014; pp. 15–33.

[8] Sweilam, N.; Tharwat, A.; Moniem, N.A. Support vector machine for diagnosis cancer disease: A comparative study. Egypt. Inform. J. 2010, 11, 81–92. [CrossRef]

[9] Nuklianggraita, T.N.; Adiwijaya, A.; Aditsania, A. On the Feature Selection of Microarray Data for Cancer Detection based on Random Forest Classifier. J. INFOTEL 2020, 12, 89–96. [CrossRef]

[10] Murugan, A.; Nair, S.H.; Kumar, K.P.S. Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers. J. Med Syst. 2019, 43, 269. [CrossRef]

[11] Aggarwal, C.C. Neural Networks and Deep Learning; Springer International Publishing: Cham, Switzerland, 2018 .

[12] Haenlein, M.; Kaplan, A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. Calif.Manag. Rev. 2019, 61, 5–14. [CrossRef].

[13] Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning; MIT Press: Cambridge, MA, USA, 2016.

[14] Russell, S.J.; Norvig, P.; Davis, E.; Edwards, D. Artificial Intelligence: A Modern Approach, 3rd ed.; Pearson: London, UK, 2016.

[15] Mcculloch, W.S.; Pitts, W.H. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 1943, 5, 115–133. [CrossRef]

[16] Russell, S.J.; Norvig, P. Artificial Intelligence: A Modern Approach, 4th ed.; Pearson: London, UK, 2020.

[17] Chiari, W.; Damayanti, R.; Harapan, H.; Puspita, K.; Saiful, S.; Rahmi, R.; Rizki, D.R.; Iqhrammullah, M. Trend of Polymer Research Related to COVID-19 Pandemic: Bibliometric Analysis. Polymers 2022, 14, 3297. [CrossRef]

[18] Patel, D.; Shah, Y.; Thakkar, N. Implementation of Artificial Intelligence Techniques for Cancer Detection.