Journal of Theoretical and Applied Information Technology
15th March 2025. Vol.103. No.5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SYSTEMATIC LITERATURE REVIEW ON BIBLIOMETRIC PROFILING ANALYSIS USING MACHINE LEARNING

ANISA IRSYAHIDAH A. RIDZUAN[1], W. M. AMIR FAZAMIN W. HAMZAH[2], MOKHAIRI MAKHTAR[3], MUMTAZIMAH MOHAMAD[4], ISMAHAFEZI ISMAIL[5], AZLIZA YACOB[6]

[1, 2, 3, 4, 5]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia

[2,3]Artificial Intelligence for Sustainability and Islamic Research Center (AISIR)

[6]Terengganu Advanced Technical Institute University College

E-mail: [1]sl4421@putra.unisza.edu.my, [2]amirfazamin@unisza.edu.my, [3]mokhairi@unisza.edu.my, [4]mumtaz@unisza.edu.my, [5]ismahafezi@unisza.edu.my, [6]azliza@tatiuc.edu.my

ABSTRACT

In the constantly expanding area of scholarly inquiry, the need for efficient instruments or tools to measure and evaluate the effectiveness of literary works cannot be overestimated. As a field of study, bibliometrics is truly informative as it measures research output in a given field, further assisting in the areas of research grants, publications and evidence for decision-making and future planning in the education sector. It is a systematic way of evaluating productivity. Hence, this research attempts to carry out a literature review regarding the application of Bibliometric Profiling Analysis in the context of Machine Learning (ML). This research involved the assessment and evaluation of the research grants, innovations, published articles, journals and proceeding papers. All materials were retrieved using online databases such as IEEE Xplore, Ingenta Connect, Sage Journal, Science Direct and Proquest. To achieve this, technologies including data mining, network analysis and natural language processing techniques were applied. From the review, the Support Vector Machine (SVM) is a machine learning algorithm frequently used in Bibliometric Profiling Analysis.

Keywords: *Bibliometrics Profiling Analysis, Machine Learning, Natural Language Processing*

1. INTRODUCTION

Bibliometrics is an analytical research technique used in library and information science that involves quantitative analysis and statistical parameters to characterize the publication trends within a certain field or numerous of literature [1]. There are several definitions: The first one was proposed by Pritchard in 1969 who defined bibliometrics as the use of mathematics and statistics on books and other media types of communications. This approach helps researchers and librarians identify the trend or pattern, influence and area of knowledge in different fields. Bibliometrics provide information about the impact of a specific paper or research article, publication frequency and authorship information for more informed decision-making in research and policy development.

Traditionally, bibliographic data is collected from known databases and other sources by scholars, whereby publications related to their subject of interest are included. It looks at the fundamental steps of data cleaning to eliminate or get rid of any duplicate values and ensure the dataset is free from any inconsistencies and contains reliable and valid information. Thus, indicators of bibliography are regarded as imperative to this study. Some examples of parameters are the h-index which measures researchers' productivity and scientific impact of researchers, the journal impact factor depicting the average frequency with which the articles of a particular journal are cited over a given period and the citation indexes defining the frequency with which one or more publications are cited. Nevertheless, it is necessary to acknowledge certain limitations of the conventional bibliometric analysis. While it provides quantitative valuable information, it may not cover qualitative aspects of specific research. According to Bornmann and Leydesdorff, bibliometric methods provide an assessment of the incomplete kind and need to be supplemented by peer review and professional evaluation to assess the research impact [2].

Given this explosion in the number of publications, there is an increasing demand for methods that would allow for efficient and thorough

processing of this immense amount of work. Complexity does not allow for a clear analysis of the patterns using traditional bibliometric methods due to the overwhelming amount of data. Machine Learning (ML) is the next possible enhancement for bibliometric analysis as it aids in analyzing huge databases with the view of predicting similar incidences in the future. Through bibliometrics, ML helps in automating the preprocessing, cleaning and acquisition of data, hence shortening the analysis time. Besides, it saves time and makes it possible to gather more compiled data from different sources with the least amount of human intervention possible. For instance, traditional techniques are often problematical in dealing with data integration from different databases and variations in references, which sometimes lead to errors and inconsistencies [3]. ML can identify connections that might be hidden in the data when compared to more traditional bibliometric analysis. However, it is important not to overlook or underestimate the limitations of previous research. For example, citing frequency and impact factors may create partial information on the research coming up with a position since they do not consider the context or quality of citations [4]. Therefore, this article examines and assesses the application of machine learning techniques in the context of bibliometrics with a focus on the aspects of data merging and the accuracy of the results obtained, as well as on recognizing complex patterns in the academic literature.

2. RELATED WORKS

In this review, several works related to ML and Bibliometric Profiling Analysis are described in more detail.

2.1 Bibliometric

As one of the main tools in research evaluation and analysis, bibliometrics has been critical in providing quantitative measures of publications' features. The term bibliometrics was first employed by Pritchard in 1969 and for bibliometric, Pritchard defined it as the statistical analysis of the properties of books and other media [1]. Over time, there has been a progressive enhancement and expansion in the application of bibliometric analysis. Leydesdorff and Milojevic further developed Pritchard's initial area characterisation in 2015. By using it to investigate the historical development of research topics, identify new trends and demonstrate the multifaceted context of scientific research, the application of bibliometric analysis has been broadened beyond mere calculation [5].

Some components of bibliometric approaches include the analysis of bibliographic data, frequency of publications, authorship and their frequency, as well as citations. Through these strategies, decision-makers, organisations and academics acquire helpful knowledge about the significance and results of their work, which results in making proper decisions. In light of digital technology development and the increase in scholarly production, bibliometrics has become an essential approach in today's scientific setting, contributing to an increased number of data points and types. Authors and organisations go on to employ bibliometric analysis to understand how scientific fields work, assess the efficiency of people and organisations, as well as make decisions beyond the academic environment. Traditional bibliometrics is defined as the quantitative analysis of bibliographic data aiming at assessing the productivity and/or the utilisation of the scholarship. In a nutshell, bibliometrics is the study of the frequency, intensity and manner of publication in scholarly literature and the details of authorship. One of the measurements most widely used in traditional bibliometrics is the citation index, which refers to the number of times a work, for example, an article, research paper or book has been cited by other works. It is commonly employed in evaluating the importance and influence of various articles within a given specialist or topic area.

The conventional applications of bibliometric measures may be applied in cases of individual researchers and organizations, such as journals and research centres, by academics and other stakeholders in organizations. Citation analysis is efficient in determining the makeup of a topic's interconnected publications, authors and journals. Also, bibliometric indices including the h-index as well as the impact factor are widely used to evaluate the relevance and quality of submitted articles. It is critical to the scientific community's decision-making regarding funding allocation and comprehensive assessments of academic institutions. Although carrying several limitations, it can be regarded as a valuable instrument of bibliometrics in the scholarly impact measurement. Several people argued that it would be misleading to completely assess the influence of the metrics derived from citations as other important aspects like societal relevance, interdisciplinary work and individuals' contribution quality might be omitted. However, traditional bibliometrics remains one of the common quantitative indicators in the system of evaluating the performance of educational institutions and can continue to contribute to further

discussions regarding scientific development and research assessment.

2.2 Machine Learning

Artificial intelligence's subdiscipline known as machine learning involves creating statistical models and algorithms that allow computer systems to improve performance at a particular task as a result of experience without the need for direct programming. It is the process of analyzing data, identifying certain patterns within the data, and being able to predict from other similar data. Machine learning algorithms are adaptive and run in an iterative loop to learn from the new information and data and then make better predictions. Through affinity, machine learning helps applied computers to reveal patterns, learn from examples and make good decisions with information input without being programmed for each scenario. It is used in several fields including speech and picture recognition, recommendation, natural language processing and predictive analysis. Supervised learning regards the training dataset as a set of input-output pairs and the algorithm aims to construct a hypothesis that would produce outputs similar to those of the training set while for unsupervised learning, the algorithm works on an unlabelled set of data to find patterns. Machine learning is a progressive, always-developing branch of science that simulates several sophisticated processes and contributes to decision-making within several spheres.

2.3 Application of Machine Learning in Bibliometric Profiling Analysis

Due to the sheer volume of work produced, there is an increasing demand for methods that could accurately and comprehensively evaluate this material as their nature bibliometric analysis methods fail to scale to the size of the problem and cannot model intricate relationships. This aspiration finds a speedy substantiation in the constantly growing body of erudite literature. It is claimed that the application of ML could significantly improve the bibliometric, meaning that it is possible to analyse patterns in big data and make prognoses of future events. In bibliometric analysis, ML makes it easier in that it automates data gathering, extraction, transformation and data scrubbing. Therefore, the collection of data from informants from several sources becomes much easier; it has less human intervention yet is very certain of a bigger sample.

Some of the previous papers have probed into the use of ML in bibliometrics in general. It is specifically about the application of such a methodology with the potential to increase the efficiency of bibliometric profiling and its accuracy, which is what attracted a lot of exciting work in this field [6][7]. To enhance the information retrieval from the given scientific literature, Belleau et al. (2008) proposed the application of machine learning approaches. Strogatz and his team used RDF triples and developed an extensive knowledge system where they highlighted the capability of machine learning in managing extensive bibliometric data [1]. In Chen's (2006) work, bibliometric and other machine-learning methods were employed to discover and map evolving patterns as well as short-lived tendencies in scientific publications. The author established that ML was applicable and effective in bibliometrics through a case study where he introduced the CiteSpace II system, which applied hybrid clustering analysis of citations [8].

Research works have also been done to understand the specific studies of related ML algorithm applicability in bibliometrics. For instance, Van Eck and Waltman (2010) designed a computer programme called VOSviewer for bibliometric mapping. The software used for the analysis and visualization of large bibliometric networks includes multidimensional scaling and clustering [9]. According to Řehůřek and Sojka (2010), using Latent Semantic Analysis (LSA), the authors managed to analyse patterns unveiled in several scientific articles. The work demonstrated an application of the ULA in identifying latent structures in bibliometric data [10].

Thus, computer science has developed significantly from 2019 to 2024, still, the techniques of creating bibliometric profiling tools are not fully elaborated. Many existing models and approaches do not fully capture new machine-learning technologies for analyzing and interpreting large-scale bibliographical data. This research seeks to solve this problem by developing an enhanced bibliometric profile model employing modern machine learning techniques like support vector machine, decision tree, k-means, hierarchical clustering, artificial neural network, genetic algorithm, k-nearest neighbor, naïve Bayes and linear regression. Thus, it becomes the purpose of this research to propose a tool for bibliometric analysis that is more accurate and versatile, independent of the area of computer science.

2.4 Profiling Analysis

For creating a more profound understanding and making sound decisions, profiling analysis is used to identify patterns, trends and traits in data. Profiling analysis allows a deeper

understanding of the composition, content and calibre of the data. This study helps researchers to obtain important characteristics of durations and assess the distribution of data and possible discrepancies. Profiling also plays its part in the creation of a clear framework for further research as it allows researchers to get familiar with the special characteristics of the data to be studied. One of the main goals of profile analysis is to obtain a better understanding of the structure of the given dataset. Statistical inferences and graphical techniques are employed to sum up and display normal propensity solutions including mainstream inclination, dispersal and regularity division. It can be useful when inferring outliers, the meaning of data skews and overviewing the overall data quality. In profiling, issues of data reliability are effectively addressed while promoting researchers' trust in other research processes and enabling them to find the most suitable approaches depending on their goals.

Also, a profile analysis readies data to perform other analyses in the future. Researchers often apply profiling in their studies to address issues such as outliers, missing numbers, formatting issues and preparation of data. If the researchers have ample knowledge of the properties of the data, they will be in a very good position to determine the methods of data transformation and data cleaning. Additionally, profiling helps the first step in determining dependencies or relations between variables, which forms the foundation for other detailed works, like correlation or predictive analysis tasks. In conclusion, the profiling analysis provides reliable and accurate research results as a whole.

2.4.1 Natural Language Processing

Jointing linguistic science, computer science and artificial intelligence, natural language processing or NLP is the interdisciplinary study that focuses on the programming of computers to understand, interpret and generate human language. Through NLP, computers can process a large amount of text material by translating it into a language comprehensible to the machine. This also means designing models and algorithms that address the issues of the vagueness of language, the ability to sense and interpret the contextual situation and sensitivity to cultural differences. NLP is applied in numerous applications involving context and meaning besides enabling good and contextual interactions with computers like sentiment analysis, language translation, chatbots and even virtual personal assistants.

The foundation of bibliometric profiling is NLP, most pronounced when machine learning is added to the process. NLP prepares the textual data of academic papers for feature extraction by ML algorithms through the steps of tokenization, stemming and stop word removal. To generate accurate profiles of researchers, institutions, or themes of the research, certain techniques are followed where basic techniques like Named Entity Recognition (NER) in which the entities like the authors, the keywords and the topics of research are identified. Moreover, it is important to note that topic modelling, which identifies important topics and themes in a large volume of scholarly documents, resulted from the interaction between NLP and ML. This enables the assessment of new trends, the status of the field and content-based publication classification. Besides, NLP contributes to the use of author profiling through the analysis of the writing behaviors such as mode of writing and grammatical usage, concerned vocabulary and topics covered. This goes a long way in helping ML algorithms create well-rounded author profiles that include specialization, areas of interest and contribution. Taking everything into consideration, NLP enhances bibliometric profiling with the help of information extraction from the materials published in scientific databases, identification of key players in a field of study, understanding the dynamics of research and predicting the trends in numerous fields of study.

2.4.2 VoSViewer

Bibliometric networks can be created and visualized with the help of the software tool called VOSviewer. It is most useful in mapping scholarly publications and journals, researchers, or keywords outlined based on citation and co-citation, bibliographic coupling and co-authorship data. One of the advantages of using it is the ability to create high-resolution graphical models that a user can later zoom in on and find out how particular data sets are related or grouped. The tool includes features for building and analyzing networks starting from data downloaded from bibliographic databases such as Scopus, Web of Science, or Pubmed, and uses distance and clustering-based mapping.

Furthermore, VOSviewer provides important benefits closely linked with the use of the software in bibliometric profiling essential for mapping and visualization of extensive bibliometric networks. It has a robust and efficient process of processing massive amounts of data and creates accurate and sophisticated maps: Citation, co-authorship and keyword co-occurrence. These visualizations are also live, meaning zooming in on specific clusters and themes and mapping between nodes such as authors, institutions, or publications.

Thanks to the basic but simple graphical user interface, VOSviewer can be easily used by Greenhorns and expert researchers in travelling through bibliometrics and the visualization of numerous bibliometric data without acquiring prior programming experience.

In bibliometric profiling, VOSviewer is popular as it provides nice and somewhat sophisticated word maps and allows for the identification of themes and research trends in large volumes of literature. Academics employ the instrument to identify key authors, institutions, or themes in certain disciplines, identify cooperative web structures and analyse the development of specific research fields. Using its highly efficient clustering algorithms, it can identify relations that are not noticeable at first employing more conventional bibliometrics approaches. This makes VOSviewer indispensable in shedding more light on the trend of research and scholarly communication in science.

For bibliometric profiling, some of the studies have, for example, applied VOSviewer. For instance, Van Eck and Waltman (2010) have shown how to use VOSviewer as a tool for mapping the scientometric field. Chen, Song and He (2021) used the tool to examine trends in renewable energy research where findings included major authors as well as themes. Donthu et al. (2021) also conducted another study aimed at identifying the evolution studies conducted during the COVID-19 pandemic and the most important articles as well as networks that work together. The study demonstrated how VOSviewer was adopted for discovering research trends, showing scholarly performance and displaying the structures of knowledge in bibliometrics studies.

2.4.3 R Studio

R Studio is a very useful tool commonly applied in bibliometric characterization to scrutinize publication histories, citations and trendy research themes and networks in multi-disciplinary contexts. Using R, researchers can use better packages such as bibliometrics and ggplot2 for efficient analysis of the data and to determine relations between authors, institutions and journals or to examine major areas or trends of interest in a specific field. R provides a high level of data handling per specification of the bibliographic data of the data sources of large-scale databases such as Scopus or Web of Science indices, conducting of descriptive statistics as well as construction of citation maps, co-authorship networks and the thematic evolution plots.

R has several advantages that make it ideal for bibliometric profiling and other applications of analysis of large data sets. One of its major advantages is that it has an open-source platform where any researcher can gain and rewrite the software as they deem fit. This openness is complemented by an extensive range of a large number of specialized packages like bibliometrics for bibliometric analysis to perform tasks as simple as handling large bibliographic datasets or as complex as performing other statistical analyses and even presenting them in the forms of sophisticated visuals. R's advantage of handling large sets of data makes it fit for bibliometric profiling, especially when working with hundreds and thousands of records in sources such as Scopus or Web of Science databases.

The use of R scripts is also repetitive and permits other researchers to easily validate their findings. As the structure of bibliographic coupling and co-citation data permits free manipulation of citation and collaboration networks corresponding to typical research themes, Ribeiro et al. demonstrated how flexibly R can be used in bibliometric profiling to uncover temporal patterns of knowledge growth in a presented subject area and produce customized visualizations. Investigators use R for these operations due to the refined statistical functions, the programmability of the environment, the community-contributed software and online R-help groups. Therefore, algorithm customization, high-speed execution and flexible application make R valuable for bibliometric profiling.

In particular, several works deal with bibliometric profiling using R. For instance, Aria and Cuccurullo (2017) unveiled the bibliometrics package and accompanied it with a use case example showing a more broad science mapping process. Donthu et al. (2021) in another study used R to map research activities related to COVID-19, which outlined the main themes and the most productive scholars. Further, Ellegaard and Wallin used R to analyze the trends of bibliometric analysis based on methods used in bibliometric analysis and showed the development of these techniques in scientific literature. All of these studies demonstrated the importance of using R in improving the process of bibliometric analysis as well as gaining a better understanding of the structure of research.

2.4.4 Comparison between VosViewer and R Studio

In the analysis of bibliometric profiling, few studies have made comparisons between R and VOSviewer, and scant attention has been paid to their capabilities and drawbacks. These comparisons primarily emphasized that both tools are suitable for bibliometric analysis but for somewhat different

purposes and concerning the users' requirements. R is generally preferred owing to its flexibility, statistical strength and versatility, which is nourished by packages such as bibliometrics. In contrast, VOSviewer is considered very user-friendly and the result is presented in easy-to-understand visualizations, even to persons with minimal computer skills.

Another article by Moral-Muñoz et al. (2020) compared bibliometric tools, such as R and VOSviewer, focusing on their data processing, mapping and visualizing capacities. The study pointed out that R offers superior statistical analysis and can be adjusted to several research practices for coding. Still, the creation of R is much more challenging than Python as users need to be familiar with programming languages and statistics. While it is not as adaptive as R, VOSviewer is good at creating network-based visualizations and ease of use. In contrast, its capacity for performing additional statistical tests is not extensive compared to the richness of R's options. Using VOSviewer's output is interactive and fully customizable, yet the programme's versatility when it comes to input and output data sources and manipulation is significantly more limited than that of R.

Petersen et al. (2022) compared CiteSpace and other 16 software, among which included VOSviewer and R; the results revealed that the former is more suitable for users who need to visualize and analyze collaboration networks or keyword co-occurrence, while the latter is more suitable for users who need to conduct more detailed bibliometric profiling, including regression analysis, clustering and others. The fact that R allows combining various bibliometric datasets and performing an array of more specific calculations predestines it for scholars in need of detailed statistics in addition to data visualization. As for the practical preference, VOSviewer is preferred by researchers who need a simple fast tool for the visualization of bibliometric data. Hence, it is particularly appreciated by users interested in easy-to-read topological maps and touch-screen graphical interfaces rather than sophisticated statistical charts. However, for those intending to perform more complex bibliometric analyses that necessarily entail the application of higher orders of statistics and tailored analytical processes, the R feature (particularly the bibliometrics packages) is highly preferred owing to the programme's flexibility at the cost of higher difficulty levels.

While there is a wide usage of the VOSviewer software for use in its appealing graphical user interface and simple and comprehensive visualizations, R on the other hand is used in cases when there is a need to considerably and further analyze the statistics and customize them. Experts in their studies decide on the choice of the tool depending on the complexity of analysis they require as well as the level of communication required with their data.

3. METHODS

The methodology concentrated on summarizing the techniques used in each prior work and choosing the most beneficial elements or procedures used in the research from previous experience to produce appropriate and accurate bibliometric profiling analysis results. This review was performed using the techniques for conducting a systematic literature review suggested by Kitchenham [11]. The process for identifying the most effective techniques or methods for bibliometric profile analysis is explained in this section. The following are the three primary steps in a systematic literature review process. The procedures conducted on each step associated with bibliometric profiling analysis are elaborated on in the next section.

3.1 Planning the Review

A systematic review is needed when all accessible information on a specific occurrence needs to be comprehensively and objectively compiled. The purpose of this systematic literature review is to support the objectives of this research, which are:

1. To provide an overview of the existing methodologies used in Bibliometric Profiling Analysis.
2. To identify the most effective ML algorithms for bibliometric profiling analysis.
3. To provide a comprehensive assessment of the data sources, preprocessing techniques and evaluation parameters for bibliometric profiling analysis.
4. To explore and uncover the potential research gaps in previous research. The creation of the research questions is the most important step in the procedure for a systematic literature review right after the research objectives are established as it drives the entire literature review process.

Kitchenham's [39] guidelines state that the population, intervention and study outcomes are the three primary components of the study criteria that should be considered while formulating research questions, which prompted the mapping of the research questions.

1. How have the methodologies for bibliometric profiling analysis evolved, and what emerging techniques were applied?
2. Which machine learning algorithms have proven to be the most effective for bibliometric profiling analysis?
3. How do different data sources and preprocessing techniques impact the accuracy and reliability of bibliometric profiling analysis using machine learning?
4. What potential gaps exist in previous research regarding Bibliometric Profiling Analysis using ML?

3.2 Conducting the Review

The design of the extraction criteria, search strategy and research selection criteria are disclosed in this section. Here are some detailed explanations for each of the covered categories.

i) Search for the Literature

Five databases were reviewed in total to ensure that relevant papers were included in this study: IEEE Explore, Sage Journal, Science Direct, Proquest, and Ingenta Connect. Table 1 contains a list of all database sources.

Table 1: The Database Sources

Database Sources	No. of Publication
IEEE Xplore	10
Sage Journal	14
Science Direct	12
Proquest	20
Ingenta Connect	11
TOTAL	**67**

The study selection criteria used to choose the best papers for this review are described as follows:

• Researchers who modeled Bibliometric Profiling Analysis using Machine Learning Algorithms.
• Review articles published and written in English.
• Studies that employed NLP.

ii) Screen for Inclusion

The primary goal was to incorporate as many publications that address research topics between 2019 and 2023. Subsequently, specific standards were established to determine whether an article qualified for inclusion in the analysis or not. A review consisting of the subsequent criteria was included in the compilation of the first batch of publications for this research.

1) Research papers on Bibliometric Analysis
2) Papers from 2019 to 2023
3) Papers published written in English

Reports that satisfied any of the subsequent requirements were not included:

1) Research without the use of data mining or machine learning methods
2) Irrelevant titles, abstracts, and keywords

The PRISMA flowchart serving as a guide during the selection process is depicted in Figure 1. There were five databases used as sources to search for the research papers, namely IEEE Explorer, Sage Journal, Science Direct, Proquest and Ingenta Connect. According to the study, 61 papers were found.

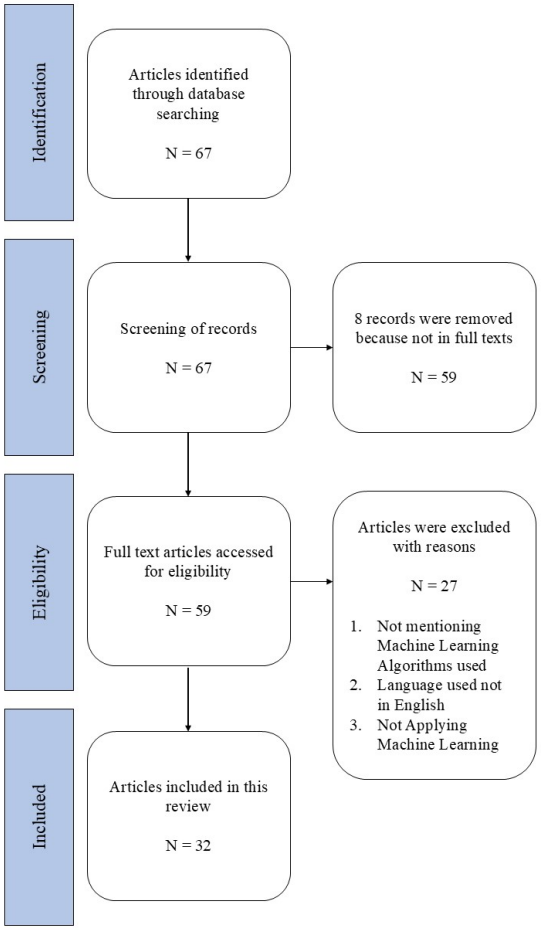

Figure 1 PRISMA Flowchart

4 Results

This section is divided into two parts, which are the Categorization of ML Algorithms and data sources as well as the findings. Each part will be explained further below.

4.1 Categorization of Machine Learning Algorithms and Data Sources

This first section delves into a comprehensive analysis of the features, methods and algorithms employed in this review. It aims to provide a thorough examination of the key elements that contributed to the study's findings and insights. Through this analysis, a deeper understanding of the methodologies and tools applied in bibliometric profiling analysis using machine learning was achieved. The inclusion of these research questions served as a foundation for conducting this study, enabling a structured and coherent exploration of the subject matter. Table 2 provides a detailed breakdown of the algorithm categories based on the completed article review, offering a clear framework for the analysis.

Table 2: Machine Learning Algorithms.

Code	Algorithms
DT	Decision Tree
SVM	Support Vector Machine
KM	K-Mean
HC	Hierarchical Clustering
ANN	Artificial Neural Network
GA	Genetic Algorithms
KNN	K-Nearest Neighbor
NB	Naïve Bayes
LR	Linear Regression

In Table 3, the frequency of algorithm usage in this research is presented and categorized. An assessment of the experimental classification model led to the identification of the most suitable and precise algorithms for bibliometric analysis when employing machine learning techniques.

Table 3 The Frequency of Algorithms Used

Published Paper	Machine Learning Algorithms								
	DT	SVM	KM	HC	ANN	GA	KNN	NB	LR
[12]		/							
[13]		/							
[14]	/			/					/
[22]									
[23]		/					/		
[25]	/	/							/
[28]								/	
[31]	/	/							
[32]					/				
[33]	/	/			/		/	/	/
[36]		/			/	/			
[37]									
[38]		/			/				
[41]							/		
[42]					/				
[46]									
[47]		/			/				
[48]		/				/			
[49]		/							
[50]							/	/	
[51]		/					/		/
[55]	/	/							
[56]									
[58]	/								/
[60]		/			/				
[62]			/						
[65]		/			/				/
[67]	/	/			/		/		
[68]	/	/					/		
[69]		/			/				
[70]		/			/				/
[71]		/						/	
Total Research Used	**8**	**20**	**1**	**1**	**11**	**2**	**7**	**4**	**7**

4.2 The Findings

Based on Table 4, the frequency of Machine Learning Algorithms was recorded. Nine algorithms were involved in previous research, namely Decision Tree (DT), Support Vector Machine (SVM), K-Means (KM), Hierarchical Clustering (HC), Artificial Neural Network (ANN), Genetic Algorithms (GA), K-Nearest Neighbor (KNN), Naïve Bayes (NB) and Logistic Regression (LR).

Table 4: Machine Learning Algorithms.

No	Machine Learning Algorithms	Count
1	Decision Tree	8
2	Support Vector Machine	20
3	K-Mean	1
4	Hierarchical Clustering	1
5	Artificial Neural Network	11
6	Genetic Algorithms	2
7	K-Nearest Neighbor	7
8	Naïve Bayes	4
9	Linear Regression	7
TOTAL		61

According to the above table, the most commonly used algorithm is the SVM, which was presented in 20 out of 61 studies and contributed roughly 32.79%. Next in line is the ANN, which accounted for about 18.03% of the total. DT comes next with 13.11% of the total. The percentage of KNN and LR was approximately 11.48%. The contributions from NB and GA were 6.58% and 3.28%, respectively.

In bibliometric profiling analysis, machine learning algorithms can be used to extract the patterns, relationships and trends in large sets of publications, citations, and research metrics. It can be noted that SVM plays the role of the most popular algorithm in bibliometric profiling. It is such an excellence of classification tasks although provided by finding the best hyperplane that divides distinct classes. For example, it is especially suitable for finding relationships in research papers, organizing research topics and forecasting future trends utilizing citation networks capable of high-dimensional data and applicable to this particular domain. SVM is popular in bibliometric profiling as it is robust in dealing with complex datasets and effective in text classification roles, also typical for bibliometric research.

KM and HC both help to group similar research papers or authors according to different attributes. KM is efficient for large datasets, while HC provides a dendrogram that helps in visualizing the clustering structure. Clustering algorithms may not always be the primary tool in bibliometric profiling, as the focus is often on classification, prediction, or regression tasks rather than unsupervised clustering.

Discussion

The features, methods and algorithms utilized in this research are considered in this section. According to the literature, there are three basic sources of success when it comes to bibliometric profiling analysis: The methodologies and techniques applied, the chosen algorithms and the quality and relevance of data [75].

(RQ:1) How have the methodologies for bibliometric profiling analysis evolved over time and what emerging techniques are being applied?

In the past, bibliometric analysis has used basic benchmark measures including citations, co-authorship analysis and journal impact factors. These approaches were mostly 'counting-driven' and aimed at estimating the efficiency of authors or institutions and determining the most 'popular' papers and journals in a particular area of interest. Bibliometric studies as a field have increasingly evolved towards methods as times changed with improvements in computation as well as data access. Co-citation analysis and co-word analysis appeared as new tools that gave a more accurate idea of the intellectual structure of academic disciplines. In the early development, bibliographic profiling was less dynamic and the data was accrued and analyzed by hands-on researchers. Sources like Scopus and Web of Science became digital, which enhanced the data availability; still, there was heavy input from human analysis. However, the application of automated tools, for example, VOSviewer and CiteSpace, was revealed to be a major shift. These tools helped researchers to work more efficiently with the data to visualize important relationships such as collaboration ones and themes in citation networks and intellectual relations that could be difficult to map otherwise. This laid the groundwork for the more advanced techniques in practice today. However, in the recent past, the integration of ML algorithms and artificial intelligence (AI) has extended bibliometric analysis further. These technologies go beyond citation indexes, utilizing such algorithms as supervised and unsupervised ML, including topic modelling, cluster analysis, as well

as network-based tools. ML makes it possible to find plug patterns and take a plunge into bigger data to identify new topics of fledging research, observe the flow of knowledge and even make forecasts as to what the next trend might look like. The shift of such practices from manual methods to AI-based analytics presents the growth of bibliometric approaches in the analysis of scientific research with modern techniques such as ML being a significant advancement. Furthermore, NLP has been recently utilized for textual feature extraction like abstracts and keywords. This makes it possible to gain an understanding of the content hence identifying underlying trends that citation methods may leave out. Optimization of the visualization of high-dimensional data has been made possible by the use of new methods which have also greatly assisted the visualization of complicated networks and themes in big data sets. The integration of ML also has the unique capability of performing real-time and large-scale bibliometric profiling of growing scientific frontiers such as the AI and biotechnology disciplines. Through the same approach, it is possible to highlight the development of methodologies throughout history as well as reveal the current state-of-the-art methods used in bibliometric studies. In conclusion, bibliometric profiling has developed from the counting of citations to all kinds of analyses based on big data and making use of the most complex of machine learning algorithms. The growth of these new kinds of research methods seems to hold out the possibility of providing a more detailed understanding of how science has evolved across various disciplines.

(RQ:2) Which machine learning algorithms have proven to be the most effective for bibliometric profiling analysis?

Advanced bibliometric profiling analysis can be exploited through the use of high-end machine learning algorithms that can process and derive insights from complex datasets. Bibliometric profiling refers to the analysis of structures and relations in documents and bibliometrics has been using less complex statistical techniques in the past. Although there have been great advancements in the way scientific articles are published and disseminated, the impact of this exponential growth has necessitated more effective, or rather scalable solutions in this line. This is the reason ML has become very important in this respect. The use of ML algorithms is beneficial for bibliometric analysis since the algorithms can process large amounts of data much faster compared to traditional methods. Therefore, they make it easier to identify relationships between authors, papers and research topics besides allowing researchers understand the trends of studies, new developing fields and the important works that should be studied. ML makes bibliometrics indispensable when it comes to classification, clustering, or detecting trends as it can do all of these automatically. It is important to specify that the supervised and unsupervised ML approaches have been used in the analyses of bibliometric profiling in recent years. These algorithms are not only restricted to the quantitative analysis of citation data but are also employed for analizing other qualitatively more diverse and intricate features, including co-authorship networks, keyword associations and topics. Other methods that have been developed include graph-based algorithms and neural networks although they provide better accuracy when used in modelling complex bibliometric data.

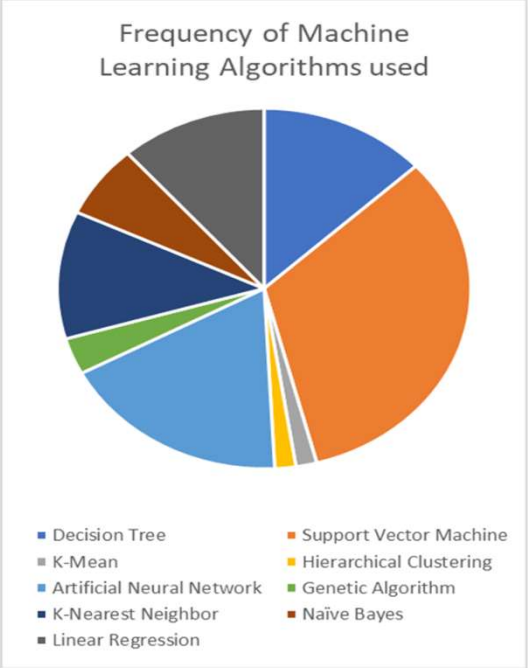

Figure 1: Pie Chart of The Frequency of Machine Learning Usage

A pie chart was developed based on Table 4 to visualise the output from the research. Nine machine learning algorithms were listed based on previous research and it turns out that SVM is the most frequently used algorithm for Bibliometric Profiling Analysis contributing 32.79% significantly more than the other algorithms. SVM are most commonly applied in bibliometric profiling analysis since they are characterised by big data capabilities and high dimensionality. In bibliometrics, since information may consist of different types of quantitative

measures and qualitative characteristics of academic literature, SVM is quite suitable for constructing hyperplanes that classify data and recognise patterns between the sets of variables. Overfitting is minimised while nonlinear relationships through kernel functions reinforce its applicability for profiling and classifying research outputs. Also, the use of SVM is effective in terms of precision necessary in bibliometric analysis where issues of classification and grouping of the research entities including authors, journals, and keywords are very sensitive. SVM provides an optimum margin that helps to distinguish between the classes, leading to the identification of trends, most influential papers and new emerging fields. In this regard, this capability is more useful in extracting insights from the large and diverse data sets inherent in bibliometric analyses.

(RQ:3) How do different data sources and preprocessing techniques impact the accuracy and reliability of bibliometric profiling analysis using machine learning?

It was also found that different data sources and data preprocessing can have a tremendous effect on the level of accuracy and robustness of bibliometric profiling analysis based on ML. Different data sources might also provide different levels of quality, data completeness and relevance, which add vulnerability to the analysis done. Data cleaning, normalisation and feature selection techniques are some of the preprocessing techniques that are so important in improving the quality of input data. Inaccurate or inconsistent preprocessing is a big risk, which leads to the wrong conclusion. On the other hand, accurate preprocessing ensure that the collected data are relevant and believable, making the performance of the machine learning model to be better. As in this research, there were five data sources visualised in the Line Chart below.

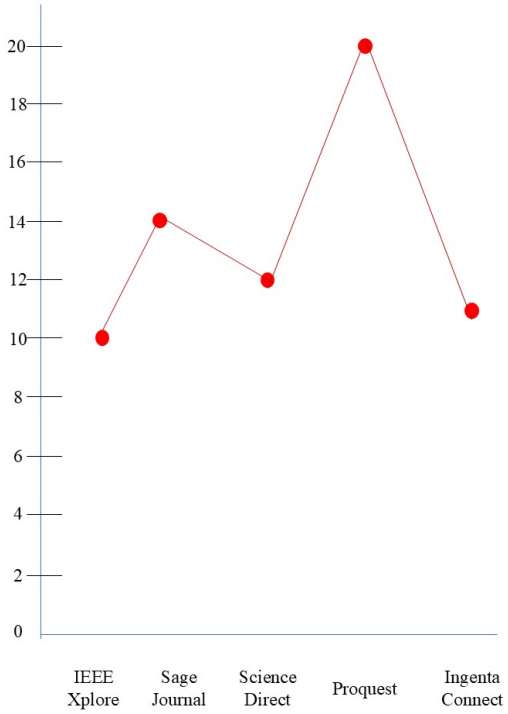

Figure 2: Line Chart of Data Sources

According to the above chart, there were five different data sources and a total of 61 journals retrieved. This is where all the above shall be cleaned and normalised to ensure that the quality of the data is good enough for analysis. Data cleaning, normalisation and de-duplication strategies play an important role in error elimination, directional conversion and removal of redundant data which in turn make the alignment of information acquired from numerous sources easier. This increases the efficiency of the ML models since it disregards noise and variability, enhancing meaningful bibliometric profiling outcomes. Preprocessing helps in preparing the data by removing noise, thereby providing true patterns and relations as well as better results for the classifiers and insights. Machine learning applications, like SVMs, have significantly transformed how academic research is analysed through profiling by improving precision and thoroughness while also boosting efficiency in the process. Unlike the metrics based on citations, ML provides a comprehensive view of researchers' profiles in multiple dimensions, enabling the capture of various activities spanning different fields. Thanks to their proficiency in managing datasets, SVMs stand out as tools for assessing research impacts and predicting future trends. This method enhances the evaluation of research influence and

collaboration connections while offering perspectives for researchers and organisations funding projects. However, data precision issues and ethical concerns continue to pose challenges that need to be addressed in the future. This study believes that SVMs have made advancements in profiling; however, there is still room for further progress and innovation in this field. By incorporating techniques like learning, one may uncover deeper connections and emerging trends. It is essential to address issues such as data privacy and algorithmic bias to ensure transparent results. Enhanced collaboration between data scientists and subject matter experts can improve the understandability and effectiveness of ML-based profiling, making it more practical for an audience within academia.

Conclusion

In this study, literature review has been carried out on bibliometric profiling analysis based on ML methods; a comparison was also made on various ways of integrating the results. Out of all the aforementioned methods, SVM was determined as the most suitable for the specified area of research. First of all, it is crucial to pinpoint the peculiarities of using SVM in bibliometric profiling analysis, which indicates its importance for the development of the study sphere. The strong preservation of classification, while at the same time being able to work with high-dimensional data, and the flexibility to address various bibliometric analyses, makes the tool highly suitable for extracting meaningful information from scholarly data. The findings derived from this study can thus form the basis for future research on other datasets and methodological extensions of conducting large-scale experiments in other sectors. Moreover, future researchers can examine the integration of the method with other techniques including other types of classifiers or new algorithms of machine learning to increase the overall level of accuracy and effectiveness of the bibliometric profiling analysis. While it could be truly said that SVMs can bring a great advantage when it comes to bibliometric profiling analysis, there are a couple of things to consider. Their learning algorithm consists of a set of parameters, which is one of their disadvantages as they are sensitive to the choice of the kernel function and parameters. Selecting an appropriate kernel to split data points in a multi-dimensional plane is another major factor of SVM. SVMs may well be more complex to apply than other types of algorithms as tuning of parameters may be quite time-consuming and professional knowledge is required to attain the best outcomes. In this case, more ML algorithms might be incorporated in the future to enhance the bibliometric profiling analysis. Techniques such as ensemble intervention, probabilistic graphical models and deep learning models such as neural networks hold the possibility of enhancing the sophistication and quality of information that can be obtained from bibliometric data. These complex algorithms can make better capabilities in terms of pattern recognition, complexity in terms of their non-linearity and existing ability to deal with all sorts of academic data.

Acknowledgment

This research was supported by the Center for Research Excellence and Incubation Management System, Universiti Sultan Zainal Abidin (UniSZA) through the University Grant with research code UniSZA/2023/SPPB (021) – RW021.

Conflict of Interest

The authors have no conflicts of interest to declare.

REFERENCES:

[1] Pritchard, A. (1969). Statistical Bibliography or Bibliometrics? Journal of Documentation, 25(4), 348-349.

[2] Bornmann, L., & Leydesdorff, L. (2014). Scientometrics in a changing research landscape. Scientometrics, 99(3), 809-815.

[3] Bornmann, L., & Marx, W. (2014). Discrepancies between planned and actual impact: evidence from scientific excellence initiatives. Research Evaluation, 23(3), 205-219.

[4] Waltman, L. (2016). A review of the literature on citation impact indicators. Journal of Informetrics, 10(2), 365-391.

[5] Leydesdorff, L., & Milojević, S. (2015). The Triple Helix of University-Industry-Government Relations. MIT Press.

[6] Nalimov, V. V., & Mulchenko, Z. M. (1971). Measurement of Science. Study of the Development of Science as an Information Process. Foreign Technology Div Wright-Patterson AFB OH.

[7] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

[8] Rousseau, R., & Zuccala, A. (2004). A classification of author co-citations: Definitions and search strategies. Journal of the American

Society for Information Science and Technology, 55(6), 513-529.
[9] Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538.
[10] Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks (pp. 45–50). Valletta, Malta: ELRA.
[11] Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. Information & Software Technology, 51(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009
[12] Xi, J., Liu, X., Wang, J., Yao, L., & Zhou, C. (2022). A Systematic Review of COVID-19 Geographical Research: Machine Learning and Bibliometric Approach. Annals of the American Association of Geographers, 113(3), 581–598. https://doi.org/10.1080/24694452.2022.2130143
[13] Andreu, L., Bigne, E., Amaro, S., & Palomo, J. (2020). Airbnb research: an analysis in tourism and hospitality journals. International Journal of Culture, Tourism and Hospitality Research, 14(1), 2–20. https://doi.org/10.1108/ijcthr-06-2019-0113
[14] Syed Asif Raza, Srikrishna Madhumohan Govindaluri, & Mohammed Khurrum Bhutta. (2022). Research themes in machine learning applications in supply chain management using bibliometric analysis tools. Benchmarking, 30(3), 834–867. https://doi.org/10.1108/bij-12-2021-0755
[15] Mulay, P., Joshi, R., & Chaudhari, A. (2020). Distributed Incremental Clustering Algorithms: A Bibliometric and Word-Cloud Review Analysis. Science & Technology Libraries, 39(3), 289–306. https://doi.org/10.1080/0194262x.2020.1775163
[16] Wan, D., Wang, R., Wei, J., Zan, Q., Shang, L., Ma, J., Yao, S., & Xu, C. (2023). Mapping knowledge landscapes and emerging trends of robotic-assisted knee arthroplasty: A bibliometric analysis. Medicine, 102(38), e35076–e35076. https://doi.org/10.1097/md.0000000000035076
[17] Wang, T., Huang, X., Zhao, L., Wang, Y., Zhang, S., Fu, X., Zhang, T., & Jiang, J. (2023). A bibliometric analysis of global publication trends on rTMS and aphasia. Medicine, 102(20), e33826-e33826. https://doi.org/10.1097/md.0000000000033826
[18] Nur, W., Teguh Haryo Sasongko, Badryia Al-Rahbi, Noorah Al-Sowayan, Asma Hayati Ahmad, Zakaria, R., Aidi Ahmi, & Othman, Z. (2023). Gene and schizophrenia in the pregenome and postgenome-wide association studies era: a bibliometric analysis and network visualization. Psychiatric Genetics, 33(2), 37–49. https://doi.org/10.1097/ypg.0000000000000336
[19] Shi, Q., & Shen, L. (2021). Orchestration capability: a bibliometric analysis. Kybernetes, 51(10), 3073–3094. https://doi.org/10.1108/k-05-2021-0386
[20] Agarwal. S., Agarwal, B., & Gupta, R. (2022). Chatbots and virtual assistants: a bibliometric analysis. Library Hi Tech, 40(4), 1013-1030. https://doi.org/10.1108/lht-09-2021-0330.
[21] Zehra Taskin, & Al, U. (2019). Natural language processing applications in library and information science. Online Information Review, 43(4), 676-690. https://doi.org/10.1108/oir-07-2018-0217.
[22] Madzik, P., Lukas Falat, Lukas Copus, & Valeri, M. (2023). Digital transformations in tourism: bibliometric literature review based on machine learning approach. European Journal of Innovation Management, 26(7), 177-205. https://doi.org/10.1108/ejim-09-2022-0531.
[23] Elnagar, A. M., Yagi, S., Ali Bou Nassif, & Salloum, S. A. (2021, February 26). Systematic Literature Review of Dialectal Arabic: Identification and Detection. ResearchGate; IEEE (Institute of Electrical and Electronics Engineers). https://www.researchgate.net/publication/349370400_Systematic_Literature_Review_of_Dialectal_Arabic_Identification_and_Detection
[24] Baek, C., & Tenzin Doleck. (2022). Educational Data Mining: A Bibliometric Analysis of an Emerging Field. IEEE Access, 10, 31289–31296. https://doi.org/10.1109/access.2022.3160457
[25] Sobreiro, P., Martinho, D., Alonso, J. G., & Berrocal, J. (2022). A SLR on Customer Dropout Prediction. IEEE Access, 10, 14529–14547. https://doi.org/10.1109/access.2022.3146397
[26] Grisales, A. M., Robledo, S., & Zuluaga, M. (2023). Topic Modeling: Perspectives From a Literature Review. IEEE Access, 11, 4066-

4078. https://doi.org/10.1109/access.2022.32322939
[27] Patricio, Soto, C., Asún, R. A., & Gutiérrez, F. (2023). Profiling support in literacy development: Use of natural language processing to identify learning needs in higher education. Assessing Writing, 58, 100787-100787. https://doi.org/10.1016/j.asw.2023.100787
[28] Zarate, D., Ball, M., Prokofieva, M., Vassilis Kostakos, & Stavropoulos, V. (2023). Identifying self-disclosed anxiety on Twitter: A natural language processing approach. Psychiatry Research, 330, 115579–115579. https://doi.org/10.1016/j.psychres.2023.115579
[29] Valdez, D., Soto-Vásquez, A. D., & Montenegro, M. S. (2023). Geospatial vaccine misinformation risk on social media: Online insights from an English/Spanish natural language processing (NLP) analysis of vaccine-related tweets. Social Science & Medicine, 339, 116365–116365. https://doi.org/10.1016/j.socscimed.2023.116365
[30] Carroll, P., Singh, B., & Mangina, E. (2024). Uncovering gender dimensions in energy policy using Natural Language Processing. Renewable & Sustainable Energy Reviews, 193, 114281-114281. https://doi.org/10.1016/j.rser.2024.114281
[31] Yuan, F., Lu, L., & Zou, Q. (2020). Analysis of gene expression profiles of lung cancer subtypes with machine learning algorithms. Biochimica et Biophysica Acta. Molecular Basis of Disease, 1866(8), 165822-165822. https://doi.org/10.1016/j.bbadis.2020.165822
[32] Hasan, M., Shahrizan Jamaludin, & Faisal, A. (2024). A critical review of machine learning algorithms in maritime, offshore, and oil & gas corrosion research: A comprehensive analysis of ANN and RF models. Ocean Engineering, 295, 116796-116796. https://doi.org/10.1016/j.oceaneng.2024.116796
[33] Barash, M., McNevin, D., Fedorenko, V., & Pavel Giverts. (2024). Machine learning applications in forensic DNA profiling: A critical review. Forensic Science International. Genetics, 69, 102994–102994. https://doi.org/10.1016/j.fsigen.2023.102994
[34] Corsi, M. P., Nham, F. H., Kassis, E., & El-Othmani, M. M. (2024). Bibliometric analysis of machine learning trends and hotspots in arthroplasty literature over 31 years. Journal of Orthopaedics, 51, 142–156. https://doi.org/10.1016/j.jor.2024.01.016
[35] Levy, A. S., Bhatia, S., Merenzon, M. A., Andryski, A. L., Rivera, C. A., Daggubati, L. C., Di, L., Shah, A. H., Komotar, R. J., & Ivan, M. E. (2024). Exploring the Landscape of Machine Learning Applications in Neurosurgery: A Bibliometric Analysis and Narrative Review of Trends and Future Directions. World Neurosurgery, 181, 108–115. https://doi.org/10.1016/j.wneu.2023.10.042
[36] Zahra Sadeqi-Arani, & Kadkhodaie, A. (2023). A bibliometric analysis of the application of machine learning methods in the petroleum industry. Results in Engineering, 20, 101518–101518. https://doi.org/10.1016/j.rineng.2023.101518
[37] Li, Y., Sha, Z., Tang, A., Goulding, K., & Liu, X. (2023). The application of machine learning to air pollution research: A bibliometric analysis. Ecotoxicology and Environmental Safety, 257, 114911–114911. https://doi.org/10.1016/j.ecoenv.2023.114911
[38] Karel Diéguez-Santana, & González-Díaz, H. (2023). Machine learning in antibacterial discovery and development: A bibliometric and network analysis of research hotspots and trends. Computers in Biology and Medicine, 155, 106638–106638. https://doi.org/10.1016/j.compbiomed.2023.106638
[39] Phoong, Y. (2022). The Bibliometric Analysis on Finite Mixture Model - Seuk Yen Phoong, Shi Ling Khek, Seuk Wai Phoong, 2022. SAGE Open. https://journals.sagepub.com/doi/full/10.1177/21582440221101039
[40] Bethlem, F. (2023). Digital Leadership: A Bibliometric Analysis - Fernanda Bethlem Tigre, Carla Curado, Paulo Lopes Henriques, 2023. Journal of Leadership & Organizational Studies. https://journals.sagepub.com/doi/full/10.1177/15480518221123132
[41] Singh, Y. (2015). A Five-Year (2015 to 2019) Analysis of Studies Focused on Breast Cancer Prediction Using Machine Learning: A Systematic Review and Bibliometric Analysis - Zakia Salod, Yashik Singh, 2020. Journal of Public Health Research. https://journals.sagepub.com/doi/full/10.4081/jphr.2020.1772
[42] Musa, H. H. (2022). Artificial Intelligence and Machine Learning in Cancer Research: A Systematic and Thematic Analysis of the Top

100 Cited Articles Indexed in Scopus Database - Ibrahim H. Musa, Lukman O. Afolabi, Ibrahim Zamit, Taha H. Musa, Hassan H. Musa, Andrew Tassang, Tosin Y. Akintunde, Wei Li, 2022. Cancer Control. https://journals.sagepub.com/doi/full/10.1177/10732748221095946

[43] Zulipikaer, L. (2023). Global research landscape on artificial intelligence in arthroplasty: A bibliometric analysis - Zhuo Li, Zulipikaer Maimaiti, Jun Fu, Ji-Ying Chen, Chi Xu, 2023. DIGITAL HEALTH. https://journals.sagepub.com/doi/full/10.1177/20552076231184048

[44] Li, M. (2020). A bibliometric analysis of 23,492 publications on rectal cancer by machine learning: basic medical research is needed - Kangtao Wang, Chenzhe Feng, Ming Li, Qian Pei, Yuqiang Li, Hong Zhu, Xiangping Song, Haiping Pei, Fengbo Tan, 2020. Therapeutic Advances in Gastroenterology. https://journals.sagepub.com/doi/full/10.1177/1756284820934594

[45] Zhenni, S. (2022). Exploring the landscape, hot topics, and trends of bariatric metabolic surgery with machine learning and bibliometric analysis - Yancheng Song, Zhenni Ni, Yi Li, Zhaopeng Li, Jian Zhang, Dong Guo, Chentong Yuan, Zhuoli Zhang, Yu Li, 2022. Therapeutic Advances in Gastrointestinal Endoscopy. https://journals.sagepub.com/doi/full/10.1177/26317745221111944

[46] Khalique, A. (2023). Bibliometric analysis on the adoption of artificial intelligence applications in the e-health sector - Abdul Khalique Shaikh, Saadat M Alhashmi, Nadia Khalique, Ahmed M. Khedr, Kaamran Raahemifar, Sadaf Bukhari, 2023. DIGITAL HEALTH. https://journals.sagepub.com/doi/full/10.1177/20552076221149296

[47] Polverino, L. (2023). Machine learning for prognostics and health management of industrial mechanical systems and equipment: A systematic literature review - Lorenzo Polverino, Raffaele Abbate, Pasquale Manco, Donato Perfetto, Francesco Caputo, Roberto Macchiaroli, Mario Caterino, 2023. International Journal of Engineering Business Management. https://journals.sagepub.com/doi/full/10.1177/18479790231186848

[48] Shaher H. Zyoud, Ahed H. Zyoud. (2021). Visualization and Mapping of Knowledge and Science Landscapes in Expert Systems With Applications Journal: A 30 Years' Bibliometric Analysis - Shaher H. Zyoud, Ahed H. Zyoud, 2021. SAGE Open. https://journals.sagepub.com/doi/full/10.1177/21582440211027574

[49] Hossein, K. (2019). Bibliometrics of sentiment analysis literature - Abdalsamad Keramatfar, Hossein Amirkhani, 2019. Journal of Information Science. https://journals.sagepub.com/doi/full/10.1177/0165551518761013

[50] Zhu, H. (2020). The Research Trends of Text Classification Studies (2000–2020): A Bibliometric Analysis - Haoran Zhu, Lei Lei, 2022. SAGE Open. https://journals.sagepub.com/doi/full/10.1177/21582440221089963

[51] Xiaojin, D. (2023). Advancements in Rumor Detection Research Based on Bibliometrics and S-curve Technology Evolution Theory - Jianbo Zhao, Huailiang Liu, Shanzhuang Zhang, Yanwei Qi, Haiping Dong, Xiaojin Zhang, Weili Zhang, 2023. SAGE Open. https://journals.sagepub.com/doi/full/10.1177/21582440231217724

[52] Joly-Chevrier, M. (2024). The State of Artificial Intelligence in Skin Cancer Publications - Maxine Joly-Chevrier, Anne Xuan-Lan Nguyen, Laurence Liang, Michael Lesko-Krleza, Philippe Lefrançois, 2024. Journal of Cutaneous Medicine and Surgery. https://journals.sagepub.com/doi/full/10.1177/12034754241229361

[53] Anupam Bonkra, Pathak, S., Kaur, A., & Mohd Asif Shah. (2024). Exploring the trend of recognizing apple leaf disease detection through machine learning: a comprehensive analysis using bibliometric techniques. Artificial Intelligence Review, 57(2). https://doi.org/10.1007/s10462-023-10628-8

[54] Martins, E., & Napoleao Verardi Galegale. (2023). Machine learning: International Journal of Innovation, 11(3), e24056–e24056. https://doi.org/10.5585/2023.24056

[55] CREDIT RATING PREDICTION USING MACHINE LEARNING TECHNIQUES: A BIBLIOMETRIC ANALYSIS - ProQuest. (2022). Proquest.com. https://doi.org/%22,

[56] Zakia Salod, & 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. Deleted Journal, 9(1),

jphr.2020.1772–jphr.2020.1772. https://doi.org/10.4081/jphr.2020.1772

[57] Nico Hananda, Azure Kamul, Harito, C., Endang Djuana, Gregorius Natanael Elwirehardja, Bens Pardamean, Gunawan, F. E., Budiman, A. S., Muhammad Asrol, A. A. N. Perwira Redi, & Pasang, T. (2023). Battery optimization by machine learning algorithms: Research gap via bibliometric analysis. E3S Web of Conferences, 388, 01020–01020. https://doi.org/10.1051/e3sconf/202338801020

[58] Umer, M., Turki Aljrees, Ullah, S., & Ali Kashif Bashir. (2023). Novel approach for quantitative and qualitative authors research profiling using feature fusion and tree-based learning approach. PeerJ. Computer Science, 9, e1752–e1752. https://doi.org/10.7717/peerj-cs.1752

[59] Umar Ali Bukar, Md Shohel Sayeed, Abdul, F., Sumendra Yogarayan, & Oluwatosin Ahmed Amodu. (2023). An exploratory bibliometric analysis of the literature on the age of information-aware unmanned aerial vehicles aided communication. Informatica, 47(7). https://doi.org/10.31449/inf.v47i7.4783

[60] Jernej Završnik, Kokol, P., Bojan Žlahtič, & Helena Blažun Vošner. (2024). Artificial Intelligence and Pediatrics: Synthetic Knowledge Synthesis. Electronics, 13(3), 512–512. https://doi.org/10.3390/electronics13030512

[61] Cardoso, M. G., Ares, E., Ferreira, L. P., & Pelaez, G. (2023/12//). Using Index Function and Artificial Intelligence to assess Sustainability: A Bibliometric analysis. International Journal of Industrial Engineering and Management, 14(4), 311-325. https://doi.org/10.24867/IJIEM-2023-4-341

[62] Bibliometric analysis for sustainable food waste using multicriteria decision - ProQuest. (2023). Proquest.com. https://doi.org/10.22034/GJESM.2023.09.SI.16

[63] Israel Edem Agbehadji, Schütte, S., Muthoni Masinde, Botai, J., & Tafadzwanashe Mabhaudhi. (2023). Climate Risks Resilience Development: A Bibliometric Analysis of Climate-Related Early Warning Systems in Southern Africa. Climate, 12(1), 3–3. https://doi.org/10.3390/cli12010003

[64] Iqbal, U., Muhammad, Zhao, J., Barthelemy, J., & Perez, P. (2023). Drones for Flood Monitoring, Mapping and Detection: A Bibliometric Review. Drones, 7(1), 32–32. https://doi.org/10.3390/drones7010032

[65] Zaidi, A. (2024). Utilisation of Deep Learning (DL) and Neural Networks (NN) Algorithms for Energy Power Generation: A Social Network and Bibliometric Analysis (2004-2022). International Journal of Energy Economics and Policy, 14(1), 172–183. https://doi.org/10.32479/ijeep.15161

[66] Zakaria, N., Sulaiman, A., Foo Siong Min, & Feizollah, A. (2023). Machine learning In the financial industry: A bibliometric approach to evidencing applications. Cogent Social Sciences, 9(2). https://doi.org/10.1080/23311886.2023.2276609

[67] Wei, Z., Liu, H., Tao, X., Pan, K., Huang, R., Ji, W., & Wang, J. (2023). Insights into the Application of Machine Learning in Industrial Risk Assessment: A Bibliometric Mapping Analysis. Sustainability, 15(8), 6965–6965. https://doi.org/10.3390/su15086965

[68] Eya Kalboussi, Ndhaief, N., & Nidhal Rezg. (2024). Last-Mile Optimization Using Neural Networks. Applied Sciences, 14(2), 787–787. https://doi.org/10.3390/app14020787

[69] Muhammad Awais, Syed, Zhang, H., Li, L., Zhang, W., Awwad, F. A., Ismail, A., M. Ijaz Khan, Raghavan, V., & Hu, J. (2023). AI and machine learning for soil analysis: an assessment of sustainable agricultural practices. Bioresources and Bioprocessing, 10(1). https://doi.org/10.1186/s40643-023-00710-y

[70] Hinge, G., Hamouda, M. A., & Mohamed, M. M. (2024). Flash Flood Susceptibility Modelling Using Soft Computing-Based Approaches: From Bibliometric to Meta-Data Analysis and Future Research Directions. Water, 16(1), 173–173. https://doi.org/10.3390/w16010173

[71] Chandra Mani Sharma, Damani, D., & Chariar, V. M. (2023). Review and content analysis of textual expressions as a marker for depressive and anxiety disorders (DAD) detection using machine learning. Discover Artificial Intelligence, 3(1). https://doi.org/10.1007/s44163-023-00090-4

[72] Jaramillo, M., Pavón, W., & Jaramillo, L. (2024). Adaptive Forecasting in Energy Consumption: A Bibliometric Analysis and Review. Data, 9(1), 13–13. https://doi.org/10.3390/data9010013

[73] Jain, A. K., None Jianchang Mao, & Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. Computer, 29(3), 31–44. https://doi.org/10.1109/2.485891

[74] Yann LeCun, Y. Bengio, & Hinton, G. E. (2015, May 28). Deep Learning. ResearchGate; Springer Nature. https://www.researchgate.net/publication/277411157_Deep_Learning

[75] Oğuzhan Öztürk, Rıdvan Kocaman, and D. K. Kanbach, "How to design bibliometric research: an overview and a framework proposal," Review of Managerial Science, Mar. 2024, doi: https://doi.org/10.1007/s11846-024-00738-0.

[76] I. Passas, "Bibliometric Analysis: The Main Steps," Encyclopedia, vol. 4, no. 2, pp. 1014–1025, Jun. 2024, doi: https://doi.org/10.3390/encyclopedia4020065.

[77] Pang, C., Wu, X., Ji, X., & Quan, V. (2020). Bibliometrics of Social Media Marketing: The Development Status and Quantitative Analysis. *International Conference on Software Engineering*. https://doi.org/10.1109/icsess49938.2020.9237686

[78] de Jesus Diaz Lara, M., Bernabe, J. G., Benitez, R. A. G., Toxqui, J. M., & Huerta, M. K. (2021). Bibliometric Analysis of the Use of the Internet of Things in Precision Agriculture. *2021 IEEE International Conference on Engineering Veracruz (ICEV)*, *20*, 1–5. https://doi.org/10.1109/icev52951.2021.9632663

[79] John, A. K., & Thomas, B. (2024). Recent Trends and Growth in E-Learning Recommender System: A Bibliometric Analysis. *2024 International Conference on Data Science and Network Security (ICDSNS)*, *1*, 1–5. https://doi.org/10.1109/icdsns62112.2024.10691270

[80] John, A. K., & Thomas, B. (2024). Recent Trends and Growth in E-Learning Recommender System: A Bibliometric Analysis. *2024 International Conference on Data Science and Network Security (ICDSNS)*, *1*, 1–5. https://doi.org/10.1109/icdsns62112.2024.10691270

[81] Mazov, N. A., & Gureyev, V. N. (2018). Modern Challenges in Bibliographic Metadata Identification. *2018 3rd Russian-Pacific Conference on Computer Technology and Applications (RPC)*, *7*, 1–4. https://doi.org/10.1109/rpc.2018.8482205

[82] Qu, H., Nordin, N. A., Teh Bor Tsong, & Feng, X. (2023). A Bibliometrics and Visual Analysis of Global Publications for Cognitive Map (1970-2022). *IEEE Access*, 1–1. https://doi.org/10.1109/access.2023.3279198

[83] Patricia, S., & Varela, J. (2022). Bibliometric Analysis of the investigation in Tax Burden. *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. https://doi.org/10.23919/cisti54924.2022.9820472

[84] Suryari Purnama, Ahmad, D., Ninda Lutfiani, Arif Darmawan, None Mardiana, Yochebed Anggraini Terah, & Azizah, N. N. (2022). *A Bibliometric Study for Blockchain Technology in Academic Journals*. https://doi.org/10.1109/iccit55355.2022.10118640