I. J. Engineering and Manufacturing, **2025, 1, 11-31**
Published Online on February 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2025.01.02

A Comprehensive Bibliometric Study on Machine Learning Based Rehabilitation and Stroke Research (1999 - 2022)

Tasfia Tahsin
Electronics and Communication Engineering Discipline, Khulna University, Khulna-9208, Bangladesh
E-mail: tasfiatahsin98@gmail.com
ORCID iD: https://orcid.org/0009-0000-0126-4858

Humayra Akter
Electronics and Communication Engineering Discipline, Khulna University, Khulna-9208, Bangladesh
E-mail: humayraakternipa180938@gmail.com
ORCID iD: https://orcid.org/0009-0003-5079-358X

Uzzal Biswas
Electronics and Communication Engineering Discipline, Khulna University, Khulna-9208, Bangladesh
E-mail: uzzal.biswas@ece.ku.ac.bd
ORCID iD: https://orcid.org/0000-0001-8489-3827

Jun Jiat Tiang
Centre for Wireless Technology (CWT), Faculty of Engineering, Multimedia University, Cyberjaya-63100, Malaysia
E-mail: jjtiang@mmu.edu.my
ORCID iD: https://orcid.org/0000-0002-1178-9356

Abdullah-Al Nahid*
Electronics and Communication Engineering Discipline, Khulna University, Khulna-9208, Bangladesh
E-mail: nahid@ece.ku.ac.bd
ORCID iD: https://orcid.org/0000-0003-2391-5767
*Correponding Author

Received: 09 August, 2024; Revised: 25 September, 2024; Accepted: 28 October, 2024; Published: 08 February, 2025

Abstract: In recent years, the rising prevalence of chronic illness has led to an increase in disability of patients. Extensive research has been done to enhance both the functional abilities as well as the quality of the affected individuals' lives. Researchers have worked on the effects of numerous scholars, keywords and countries of these specific fields. However, a few state-of-the-art bibliometric analyses have been done in this research to reduce the quantitative aspects of the vast research fields of rehabilitation. We have precisely selected 427 core papers from the Web of Science database spanning from 1999 to 2022 where Machine Learning (ML) or Deep Learning (DL) is used in the rehabilitation field. Consequently, our analysis focuses on citation patterns, trend analysis and collaborations between countries or influential keywords offering a detailed overview of global trends in this interdisciplinary domain. Additionally, we visualize the research trends of various authors and countries which provide invaluable insights into research impact as well as collaboration networks. Overall, this paper aims to shape the evolving field of rehabilitation by providing in depth analysis of the citation landscape, key researchers, and international collaborations.

Index Terms: Rehabilitation, Stroke, Machine Learning, Deep Learning, Bibliometric

1. Introduction

The evolution of rehabilitation practices that include massage of different body parts exercises or hydrotherapy in ancient civilizations, strengthened the foundation for current stroke rehabilitation techniques. In addition, the advancements of the processes of physical healing, prosthetics, and psychological care developed during the Industrial Revolution and the World Wars have noticeably influenced the comprehensive approaches used in stroke rehabilitation today[1].

Bibliometric analysis of the rehabilitation field implies the evaluation of the relative importance of scientific production in a particular field. It allows new scholars or researchers to acquire some informative knowledge within a certain field as well as build up an interdisciplinary collaborating network[2]. It also represents the geographical distribution of any work in a particular field. Moreover, the international cooperation of various authors, affiliations, author's countries, and so on can be expressed with the help of bibliometric analysis. In recent years, some bibliometric analysis-based papers have been published among which rehabilitation using virtual reality by Youliang et al. has gained much attention. The main aim of the study was to conduct a retrospective analysis of the research articles regarding rehabilitation medicine with the help of virtual reality technology from 1996 to 2015. Moreover, they categorized the most influential journals with their total publications as well as the 160 institutions. Apart from these, this article emphasized the prolific authors, high LCS (Local Citation Scores) papers and their journals, top subject areas, article cluster analysis, and so on[3].

Xiali Xue et al. analysed the international research data from 2010 to 2020 of Web of Science which was about global trends and hotspots in research on the field of rehabilitation robots. They included 3194 articles from 82 countries and classified more than 50 articles into 25 categories. They used CiteSpace software to analyse and represent different relationships and cooperations. Moreover, they categorized the 10 influential journals with their number of total publications and citations. The top 10 highly prolific authors are also represented with their countries and number of publications with citations. They used an emergent word detection algorithm and frontier analysis for this purpose[4]. Post-stroke rehabilitation has gained much attention nowadays as only 10% of stroke patients fully recovered and most of them need rehabilitation to come back to their normal life. Xiaodong Feng et al. quantitatively analysed the research progress in the rehabilitation treatment of stroke patients. They used Web of Science databases from 2003 to 2013 and indexed 2654 clinical trial investigations. They selected the outcomes of the trials by overall trial numbers, distribution of output by trial regions, country, and institutions, number of received citations, and so on[5].

Meanwhile, some of the authors analysed and researched the bibliometric analysis of rehabilitation, but none of them represented a complete package of the authors, country, affiliation, collaboration, and correlation in the field of rehabilitation. Thus, in this paper, prolific authors, their countries, institutions, and keywords will be sequentially represented as well as the collaboration network between them. Moreover, machine learning or deep learning for rehabilitation of the post-stroke or elderly patients, and clinical exercise of the patients are represented here. A review of machine learning or deep learning-based rehabilitation analysis with sensors is quite new. The major contributions of the paper are as follows,

1. Visualizing the overall scenario of the prolific scholars, institutions, affiliations, and keywords at a glance and the collaboration between them.
2. Summarizing globally renowned scholars, their research trends in the rehabilitation field to guide future research fellows.
3. Analysing the benefits and future research scopes of using machine learning or deep learning models to estimate rehabilitation.

The second chapter demonstrates the methodology of the research which illustrates the data collection as well as data evaluation. Furthermore, the third chapter represents the results of the analysis in terms of keywords, citations, collaboration and trends. Finally, the last chapter concludes with the probable future works of this research.

2. Methodology

In the realm of rehabilitation and stroke research, the proliferation of studies regarding machine learning or deep learning methodologies has made manual review which is a challenging task. To mitigate the challenges regarding manual analysis of the researches as well as the examination of research papers, two data mining software tools, Biblioshiny and VOSviewer, have been employed. These tools not only lessen the burdens associated with manual labour but also simplify the systematic and efficient analysis of literature in this domain. The whole process of selecting the keywords to analyze the papers is represented stepwise in Fig. 1 through a block diagram in.

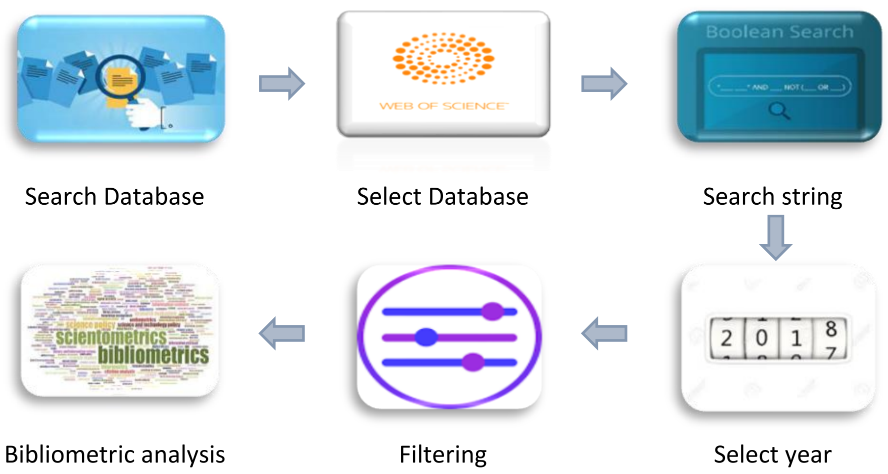


Fig. 1. Research methodology block diagram

2.1. Bibliographic data collection

Different data sources provide various kinds of data that can be used for bibliometric analysis. Web of Science, Scopus, PubMed and Google Scholar are available but we have extracted our data from Web of Science. The search string that we used to retrieve our required dataset is "AB= (rehabilitation AND (stroke OR health or clinical or exercise or elderly care) AND Sensor AND (Machine Learning OR Deep Learning))" and we only have datasets from 1999 to 2022. We have 427 documents from 275 sources after excluding some criteria such as English language and review articles. We have used Biblioshiny and VOSviewer software to analyze and plot the graphs and figures. We have also used tables and figures to visualize the graphical as well as theoretical representations. Table 1 below represents Primary information about the documents, authors, and citations with the summary of the dataset[6].

Table 1. key information about data and documents

Data information	
Period	1999:2022
Number of Sources	275
Documents	427
Average years from publication	3.71
Mean citations per document	9.686
Mean citations per year per doc	1.844
References	15697

PRISMA flow diagram is detailing the systematic process of identifying, screening, and including studies on rehabilitation-related research using machine learning and sensor data from 1999 to 2022, resulting in 427 documents from 275 sources after applying exclusion criteria which has been presented in the Fig. 2. This rigorous process ensures transparency and replicability in the analysis of rehabilitation-related research using machine learning and sensor data.

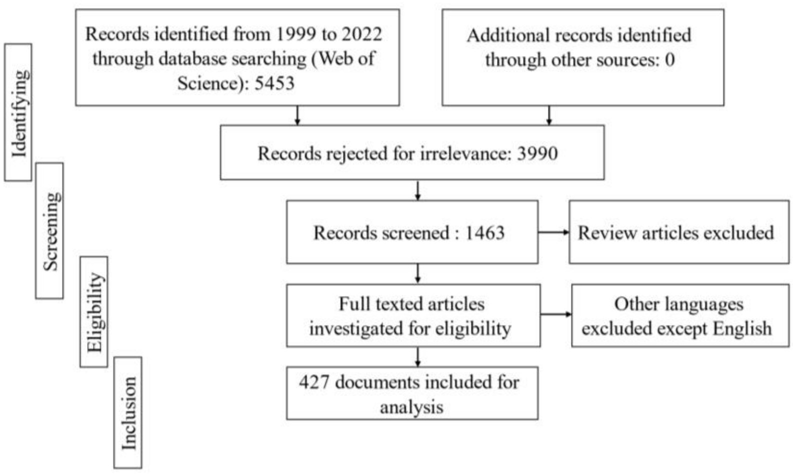


Fig. 2. PRISMA framework for the systematic process

2.2. Bibliographic data evaluation

This section investigates citation, trend analysis as well as network and collaboration analysis based on the authors, countries and keywords. To effectively manage a multitude of data for analysing diverse correlations and displaying various trends, certain parameters are utilized namely citation, authors, keyword and affiliation.

- Citation

Analyse the cited references as well as the local and global citations one of the aims to evaluate the results of the study. Moreover, the number of citations of the countries or sources under a time period is also illustrated in terms of citation. Citation analysis mainly focuses on the citation trends and most cited topics, countries or affiliations to ease the research topics and ideas for a new researcher in this estimated field.

- Authors

This section focuses on authors as well as their influences on the collected articles for rehabilitation research. Author analysis mainly works to find out the significance of features related to authors. A total of 2026 authors are involved in this rehabilitation research work among which only 13 are sole-authored. The rest of the 2013 documents are multi-authored documents. The most relevant author representation was a part of the citation analysis that finds the most qualified and skilled authors in this rehabilitation field. Author qualification depends on the number of documents published by an author per year as well as the number of citations he receives per year. One author may get fewer citations while publishing more specialized documents in a certain field than another author with fewer publications[7].

- Keywords

Keywords play a vital role when considering any research paper, to sum up, the fundamental issues they represent. So, we analyze the keywords in accordance with the frequency of occurrence, growth as well as thematic evolution. In this case, words and their co-occurrences are the units of analysis.

Generally, Author's Keywords (DE) and Keywords Plus (ID) are the two types of keywords that we use to represent Bibliometric analysis. The first one is the Author's keywords that the author uses in their research papers. On the other hand, keywords are generated in Keyword Plus that use the Bibliometrics software and algorithms. We have 856 Keyword Plus (ID) and 1264 Author's keywords in our data. Though the co-citation or bibliographic coupling analyzes their cited papers whereas keyword analysis represents the most used or important keywords of the content. Moreover, we analyze keywords based on author keywords but we can also find keywords from abstract, introduction, conclusion, and so on. Subsequently, we can focus on some selective keywords on which we can go for future research in this field[8].

- Collaboration and Bibliographic Coupling

Bibliographic collaboration involves the study of co-authorship relationships among researchers. This is often examined using co-authorship networks, where the nodes signify authors and the edges represent their collaborative works. Conversely, bibliographic coupling happens when two documents cite the same third document in their references. This approach measures the relatedness of documents based on their shared references, indicating similarities in subject matter and emerging research trends.

3. Scholarly Impact Analysis

To analyze the impact of numerous prolific authors, countries or sources and their collaborations as well as trends are discussed in the section so that investigating the related researches become easier and more convenient.

3.1 Citation analysis

The Bibliometrix R package links the parameters and represents the rates of various interconnected fields of the human action recognition dataset. Citation analysis according to cumulative occurrences through years of articles is also discussed in the section.

Here in Fig. 3, the mean total citation per year is expressed for 1999 to 2022. The horizontal axis represents the timespan whereas, the vertical one indicates the number of mean total citations for the certain timespan. If we analyze the mean total citation per year, the rate of change is very less till 2006 and then suddenly increases in 2009. Then the number gradually decreased for the rest of the years and reached 1.5 in 2022.

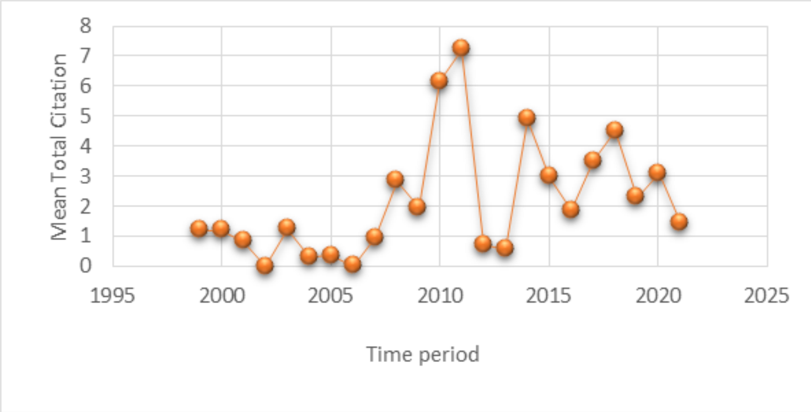


Fig. 3. Mean TC per year

Fig. 4 expresses the 20 most locally cited sources with their citations. Among the sources, SENSORS-BASEL has the topmost local citations and STROKE got the least citations.

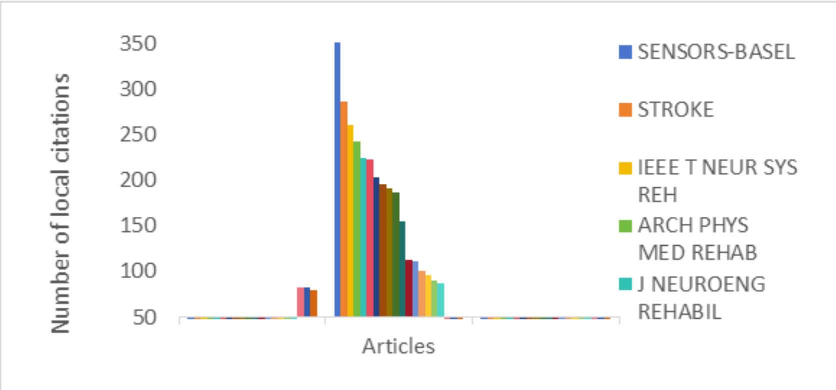


Fig. 4. Most locally cited sources

With the help of citation analysis, frequency, patterns, and graphs of the citations of any articles or authors come out. Local citations (LC) and global citations (GC) are the two types of citations an article can get. If two authors or documents of the same collection cite each other, the number of citations is called local citations. Among all the documents, the 20 most impactful documents with local as well as global citations are demonstrated here. Here, SENSORS-BASEL got 9 local citations whereas EEE Transactions on Biomedical Engineering has 8 local citations. The rest of the documents are following and the local citation is gradually decreasing. On the other hand, global citations represent the citations received from the documents or authors out of the collection but indexed in a bibliographic database. Here in Table 2, local and global citations with their ratio are presented from which normalized local and global ratios can be extracted. SENSORS-BASEL has the highest number of local citations but that doesn't mean that it has the topmost global citations. Rather than the SENSORS-BASEL, Expert Systems with Applications got 241 global citations which is the highest among them. The relative citation ratio or Local citation to global citation ratio (LC/GC) represents the ratio of the two citations.

Table 2. Local citation and global citation-related information

Document /year	DOI	LC	GC	LC/GC Ratio (%)	Normalized Local Citations	Normalized Global Citations
OTTEN P, 2015, SENSORS-BASEL	10.3390/s150820097	9	43	20.93	7.36	2.04
PANWAR M, 2019, IEEE T BIO-MED ENG	10.1109/TBME.2019.2899927	8	34	23.53	21.91	4.9
NWEKE HF, 2018, EXPERT SYST APPL	10.1016/j.eswa.2018.03.056	7	241	2.9	9.84	13.34
YANG G, 2018, IEEE J TRANSL ENG HE	10.1109/JTEHM.2018.2822681	7	60	11.67	9.84	3.32
LIN WY, 2018, INT J MED INFORM	10.1016/j.ijmedinf.2018.01.002	6	27	22.22	8.44	1.49
LAU HY, 2009, HUM MOVEMENT SCI	10.1016/j.humov.2008.12.003	5	47	10.64	2.86	1.84
GOGGINS OM, 2014, J NEUROENG REHAB	10.1186/1743-0003-11-158	5	63	7.94	3.75	1.6

SOEKADAR SR, 2015, NEUROBIOL DIS	10.1016/j.nbd.2014.11.025	5	151	3.31	4.09	7.17
YU L, 2016, COMPUT METH PROG BIO	10.1016/j.cmpb.2016.02.012	5	50	10	8.5	4.43
IWAMOTO Y, 2020, J STROKE CEREBROVASC	10.1016/j.jstrokecerebrovasdis.2020.105332	5	8	62.5	14.5	1.28
BOCHNIEWICZ EM, 2017, J STROKE CEREBROVASC	10.1016/j.jstrokecerebrovasdis.2017.07.004	4	20	20	9.2	1.13
LIAO YL, 2020, IEEE T NEUR SYS REH	10.1109/TNSRE.2020.2966249	4	18	22.22	11.6	2.89
DOBKIN BH, 2011, NEURO REHAB NEURAL RE	10.1177/1545968311425908	3	174	1.72	2.5	2.18
SALE P, 2018, J STROKE CEREBROVASC	10.1016/j.jstrokecerebrovasdis.2018.06.021	3	8	37.5	4.22	0.44
CHAE SH, 2020, JMIR MHEALTH UHEALTH	10.2196/17216	3	13	23.08	8.7	2.09
IMURA T, 2021, J STROKE CEREBROVASC-a-b	10.1016/j.jstrokecerebrovasdis.2021.105636	3	3	100	28.64	2.09
AMIRABDOLLAHIAN F, 2001, ASSIST TECHN RES SER		2	16	12.5	2	0.89
ZHU M, 2007, J CLIN EPIDEMIOL	10.1016/j.jclinepi.2007.06.001	2	23	8.7	2.67	1.59
AKDOGAN E, 2009, J INTELL MANUF	10.1007/s10845-008-0225-y	2	47	4.26	1.14	1.84
DOBKIN BH, 2011, STROKE	10.1161/STROKEAHA.110.611095	2	63	3.17	1.67	0.79

Fig. 5 shows the 20 most cited countries with the citations they got within a certain time. The USA has 886 citations and China has 570 citations which are the two highest citations among all the countries.

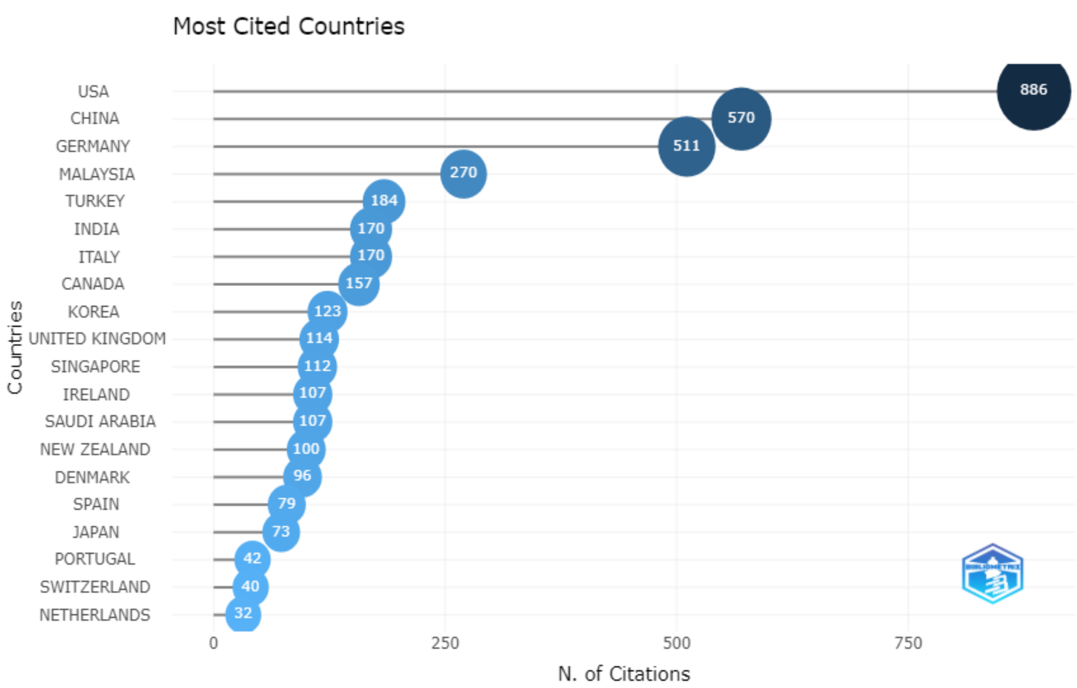


Fig. 5. Most cited countries

The first three countries got more than 500 citations and the rest of them got less than 300 citations. Portugal, Switzerland, and Netherlands got less than 50 citations which is the least among the countries.

Reference publication year spectroscopy (RPYS) basically depends on the frequency of the references that are cited in the publications of this rehabilitation field. The annual distribution of the cited references for 30 years which is from 1990 to 2020 is demonstrated here. To analyze the historical root of a specific research field or topic, this RPYS is another important tool. We can assume from the Fig. 6 that, cited reference is only visible in 1994 before 1999 and then gradually increased after 1999. Moreover, this figure indicates two peaks of 10 and 14 references in 2010 and 2012. The figure has a linear increment from 2013 and reaches its peak in 2015 with more than 30 citations. Then it started to decrease from 2015 to 2020 and received less than 10 citations again.

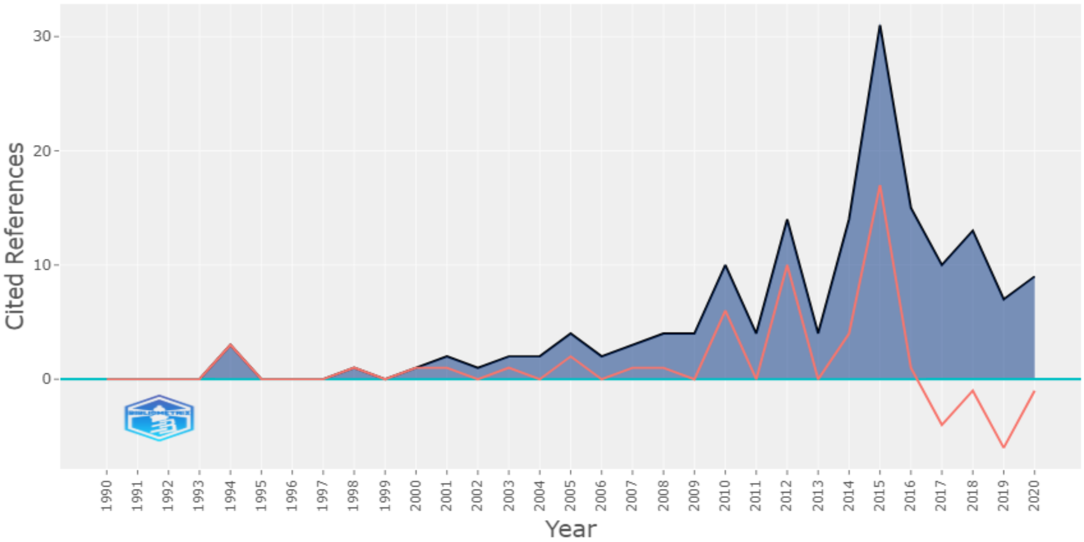


Fig. 6. Reference publication year spectroscopy

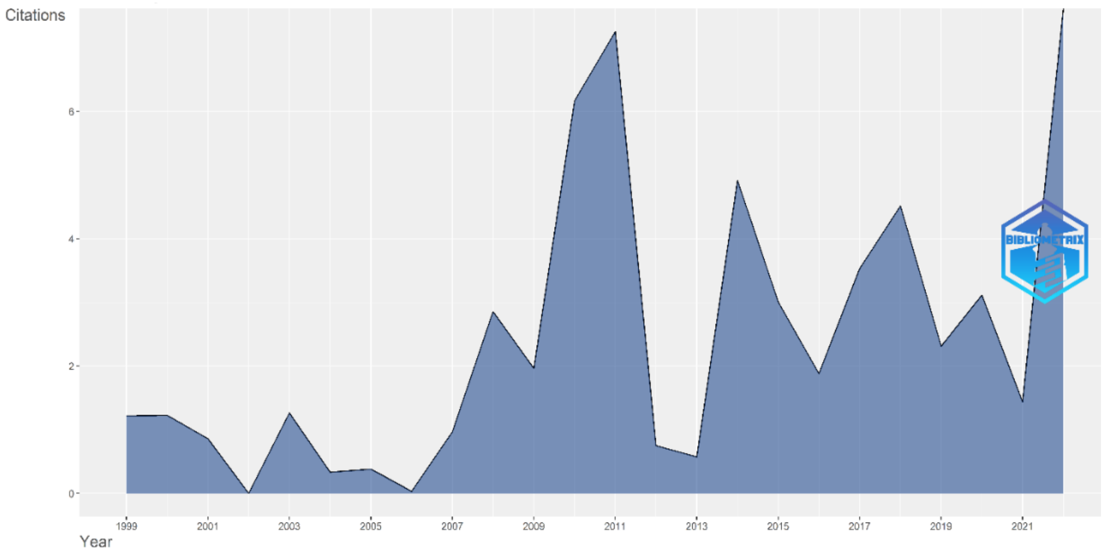


Fig. 7. Average article citation per year

Average article citation per year is the ratio of the total number of citations to the number of years the author or journal has been publishing papers. This metric is a beneficial tool to measure the yearly impact of any author or source. Here Fig. 7 implies the average citation per year from 1999 to 2021 and represents them with a useful figure. The average citation was not too high till 2007 and it increased rapidly and got more than 7 average citations in 2011. Though average citations rapidly fell apart in 2012 that was less than 1 citation per year. However, the average citation was quite good from 2013 to 2021 which was between 3 to 5 average citations per year. Finally, the average citation reaches its peak in 2022 when the number of average citations is almost 8 citations per year.

3.2 *Keyword evaluation*

The main aim of this keyword analysis is to find out the most trending topics and authors of this rehabilitation field. Last but not least, we can predict the future trajectories of this certain field from this keyword clustering.

(a) (b)

Fig. 8. Word cloud of authors' keyword(a) and keyword plus(b)

Here, Fig. 8(a) displays the word tree for author keywords so we can differentiate it from Fig. 8(b) where keyword plus is used to make a word cloud. Here machine learning, rehabilitation, and deep learning are the most used words with 124, 64, and 49 frequencies whereas sensor and classification have the lowest occurrence rate in the rehabilitation field[9]. So, we can observe from Table 3 that the author's keywords have a higher frequency than the keywords of Keyword Plus.

Table 3. Authors keyword vs. keyword plus

Author's keyword		Keyword plus	
Keyword	Occurrence	Keyword	Occurrence
machine learning	124	rehabilitation	36
rehabilitation	64	stroke	35
deep learning	49	classification	34
stroke	47	recovery	34
artificial intelligence	14	system	26
stroke rehabilitation	14	therapy	20
wearable sensors	14	recognition	19
classification	12	reliability	18
feature extraction	12	gait	16
sensors	12	care	15

The word cloud of Fig. 8(b) represents 50 authors' keywords of different sizes in accordance with their frequency of appearance in the rehabilitation field. Rehabilitation, stroke, and classification are the most dominant words which are in the middle and large whereas reliability, gait, and care are the least important words. Rehabilitation is the most repetitive word with 36 frequencies among them and then comes stroke, classification, and so on[10].

One of the most popular methods for keyword analysis is discussed in Fig. 9 where authors' keywords and their occurrences are demonstrated along with the number of clearances. So, one can easily find out the most relevant keywords or most frequently used keywords in this field.

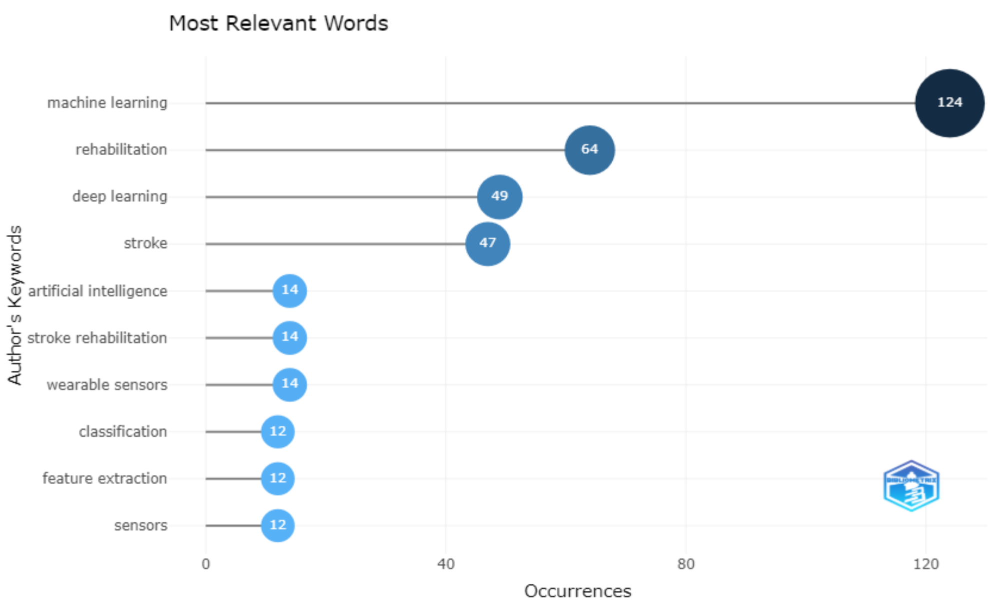


Fig. 9. Most relevant words

The author's keywords indicate the used keywords of the author. We can visualize the topmost impactful words of the rehabilitation field with this chart. The horizontal axis contains the frequency of occurrence of the keywords and the horizontal axis represents the keywords. As we can see machine learning is the most relevant keyword and then gradually comes rehabilitation, deep learning, stroke, and so on. On the contrary, sensors have less impact on this rehabilitation field with 12 occurrences.

The purpose of the conceptual structure map is to perform multiple correspondence analysis of dominant words. Moreover, K-means clustering helps to find out clusters of similar conceptual contents.

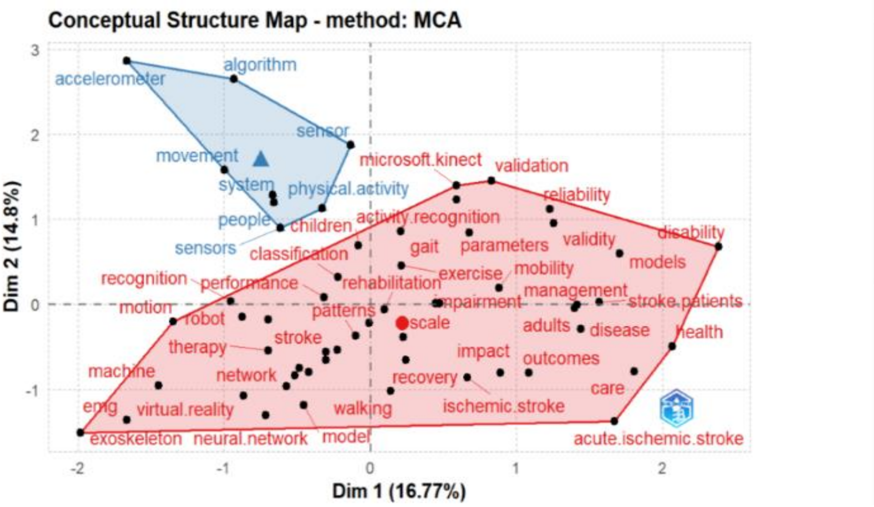


Fig. 10. Conceptual structure map

K-means clustering ensures the two different clusters here in Fig. 10 with the help of the unsupervised classification method. Here we put every word according to its dim 1 and dim 2 values. Red and blue are two clusters that represent the[11] connection between the words used in this field. MCA method graphically as well as numerically analysis our rehabilitation-related data. So we can find out latent variables from the interdependent variables of the clusters[11].

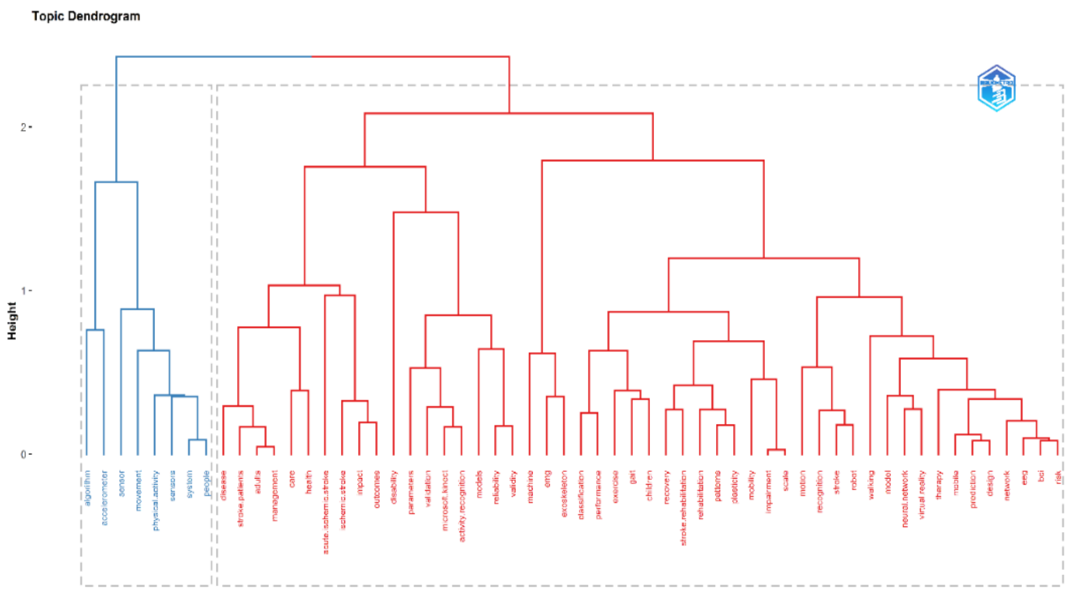


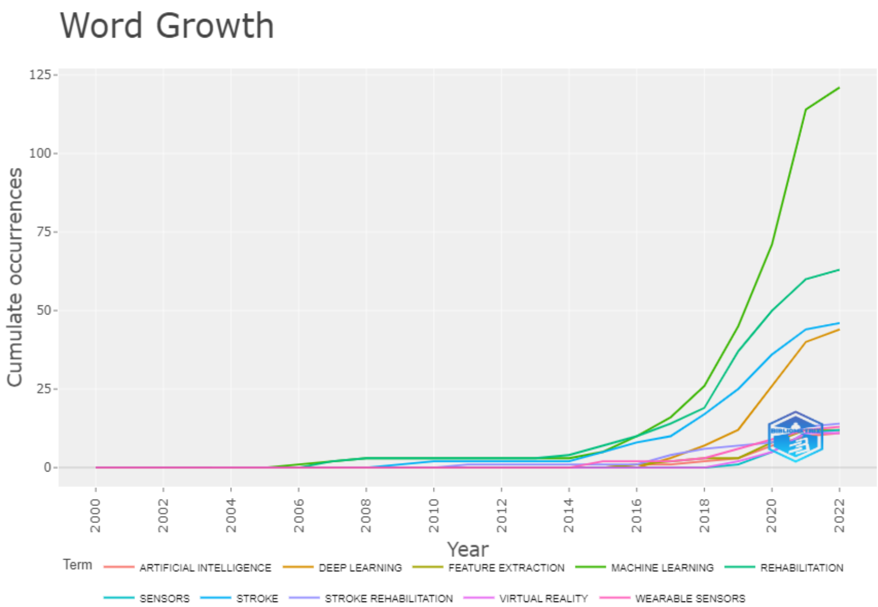
Fig. 11. Dendrogram map

This dendrogram represents a link between keywords with the help of hierarchical clustering. We can visualize the interconnection of different words and related topics among clusters. In Fig. 11, topics are classified into two different colours blue and green and the topics are close enough. Moreover, the topics are divided into clusters and sub-clusters. We can set a threshold to split the dendrogram into several clusters. Instead of presenting the degree of clusters, the dendrogram represents several clusters. We consider the clustered keywords as different classes and so in accordance with the similarity of degrees, we can combine two or more clusters[12].

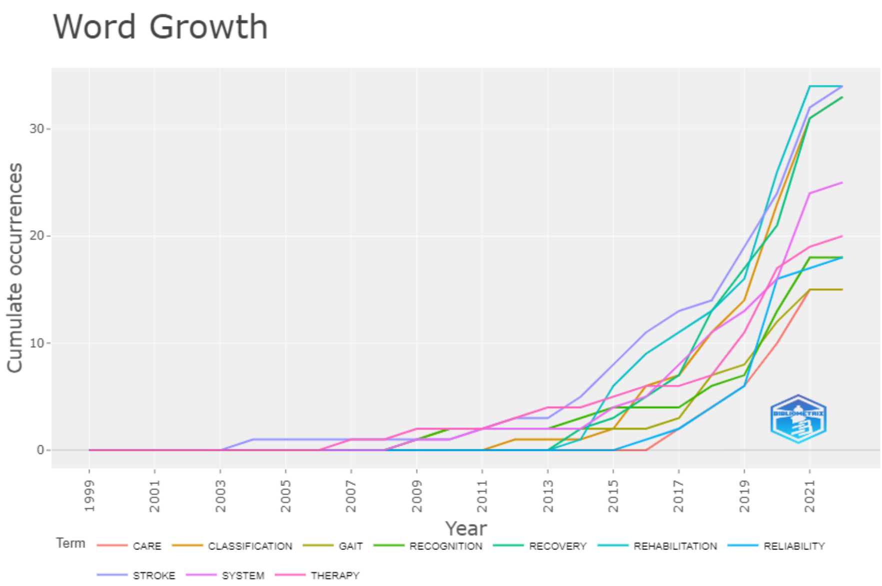
According to the software analysis, this tree diagram divides various interconnections among elements. The height of the coordination line has an impact on the different groups as well as clusters. So, every cluster gets merged with higher degrees of similarity and thus the process is repeated in this rehabilitation dendrogram.

Fig. 12. Treemap of author's keyword

In Fig. 12, a tree map of keywords represents how frequently a keyword is used and its dominance over a certain field. So, in our rehabilitation field, the keywords rehabilitation, stroke, and classification have a great impact of 8% and so they are in the first column. The following keywords are recovery, system, and therapy having 8%, 6%, and 4% frequency impact, and so on. Thus, the top 30 keywords have certain rows that determine their effectiveness and the percentage demonstrates the frequency of occurrences. Exercise, walking, EEG, and motion have comparatively lowest contribution in the rehabilitation field as well as having 3% and 2% frequencies[13], [14].



(a)



(b)

Fig. 13. Cumulative occurrences of authors keyword(a) and keyword plus(b)

We can visualize the growth rate of the Author's keywords for a certain time with Fig. 13(a). Here we have taken the rate of growth for 22 years with the cumulative occurrence of the top 10 keywords. The keywords started to appear in this rehabilitation research landscape in 2014 and a larger growth has been observed since 2018. Machine learning grew more rapidly than the other ones exactly from 2018 and gradually rehabilitation, stroke, and deep learning are followed. The faster rate started to slow down in 2021. The rest of the keywords have a slow growth rate till 2019 and increased a little bit after that.

On the other hand, we can compare the growth rate of keyword plus in Fig. 13(b). If we analyze the rate of growth, most of them appeared rapidly from 2013 and the second wave of growth is observed in 2019. Strokes started growing faster in 2013 and rehabilitation replaced them after 2019. Recognition, care, and gait have had a slower rate of growth in the recent few years. All the keywords of keyword plus grew simultaneously in keyword plus whereas we can see that some of the keywords have a fast growth rate in the author's keyword[7].

We can visualize the growth rate of the Author's keywords for a certain time from the figure too. Here we have taken the rate of growth for 22 years with the cumulative occurrence of the top 10 keywords. The keywords started to appear in this rehabilitation research landscape in 2014 and a larger growth has been observed since 2018. Machine learning grew more rapidly than the other ones exactly from 2018 and gradually rehabilitation, stroke, and deep learning followed. The faster rate started to slow down from 2021Rest of the keywords had a slow growth rate till 2019 and increased a little bit after that.

The thematic map in Fig. 14 demonstrates four different types of theme analysis according to the two axes of density and centrality. The horizontal and vertical axis represents density and centrality in the case of four different quadrants. Density generally describes the strength of the relationship between keywords and the other property centrality depicts the degree of correlation among various trending topics.

Hence, these two properties make sure if a topic is well developed or not as well as its importance. If a thematic node has a higher amount of connectivity with the other nodes, its centrality, as well as importance, is high as well. Therefore, the node will stay in an essential position in the thematic map. On the contrary, density defines the opportunity to develop any theme or its sustainability of them. m

So, we divided the map into four quadrants namely Q1, Q2, Q3, and Q4 with different values of density and centrality. The right-side upper quadrate contains the developed elements with the highest density as well as centrality. Just like the exoskeleton, motion and virtual reality have importance above average. The other two nodes containing rehabilitation, stroke, recovery and care, health, and disease are well developed and have immense power to shape the rehabilitation research field.

Exercise, impact, and physical activity are in the Q4 and so their importance is high in this field as they are the basics. Nodes of the Q2 gained developed interconnections among other nodes but the development ratio is still not so high for them. So they have potential but need more connection with rehabilitation. Last but not least, the nodes of Q3 need more attention and effort to develop as they are also very basic and essential[11].

Among the 275 sources of this work, Fig. 15 explains the growth of the most impactful 7 sources. From 1999 to 2021 timespan, the cumulative occurrence of the sources is represented in this figure. We can say that the sources were introduced in 2013 and started to get significant in 2018. "IEEE Access" initiated its journey in 2015 but we can see a linear growth of this second top journal from 2018. The most relevant journal "Sensors" started publishing articles in 2014 and the cumulative occurrences were below 5 till 2019. After that, it reached the topmost position linearly where cumulative occurrences are up to 20 in 2021. The other journals are following and most of them started to grow up linearly after 2018. “Frontiers in neuroscience”, “Frontiers in neurology” and so on are following with a linear growth since 2021.

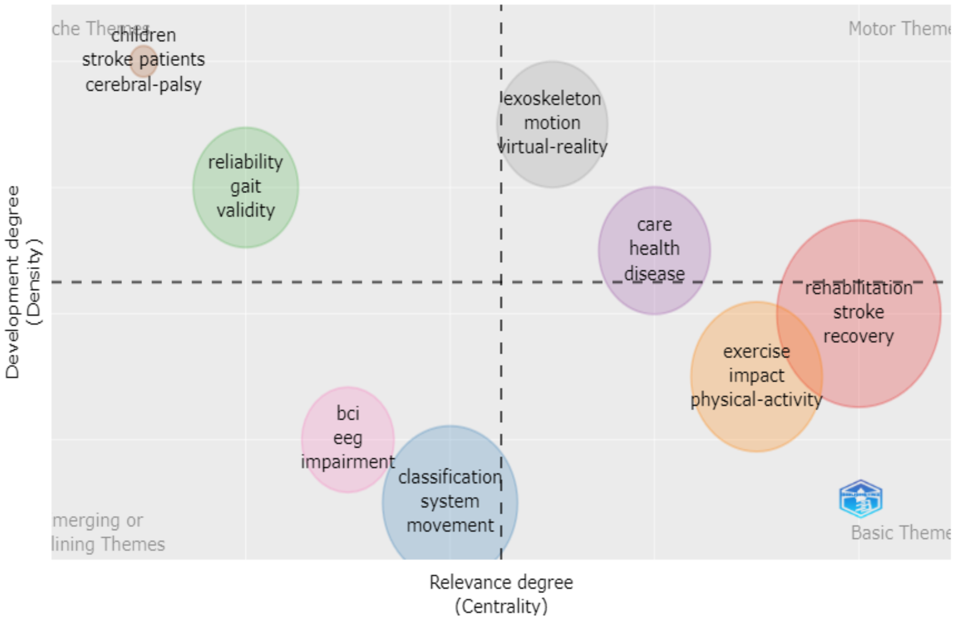


Fig. 14. Thematic map of keywords

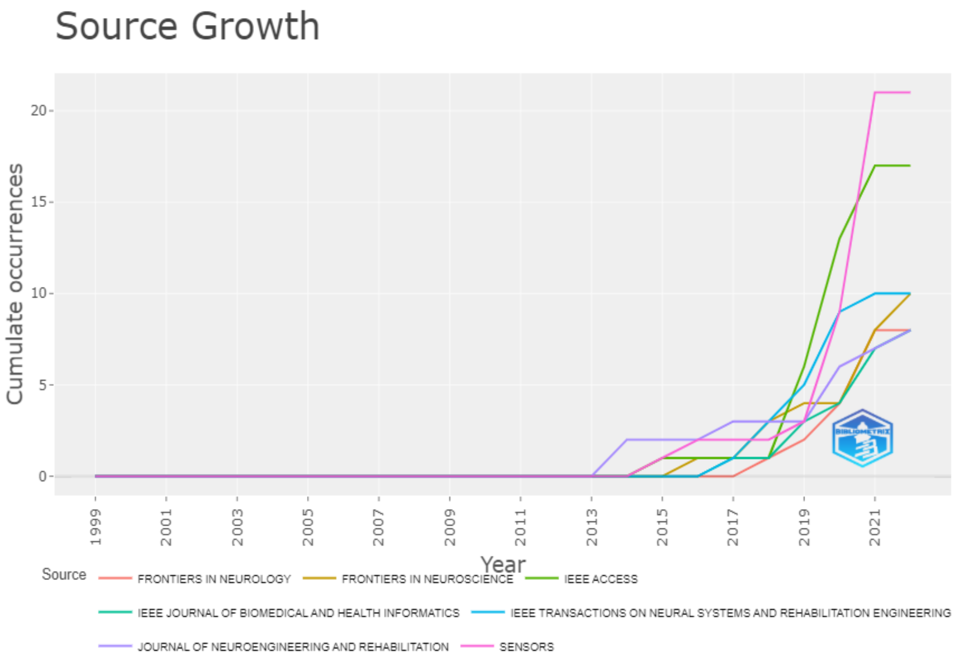


Fig. 15. Source growth in terms of cumulative occurrences

3.3 Bibliographic collaboration and coupling

This section presents a collaborative analysis of review papers by visually mapping the relationships and collaborations between authors, institutions, and countries.

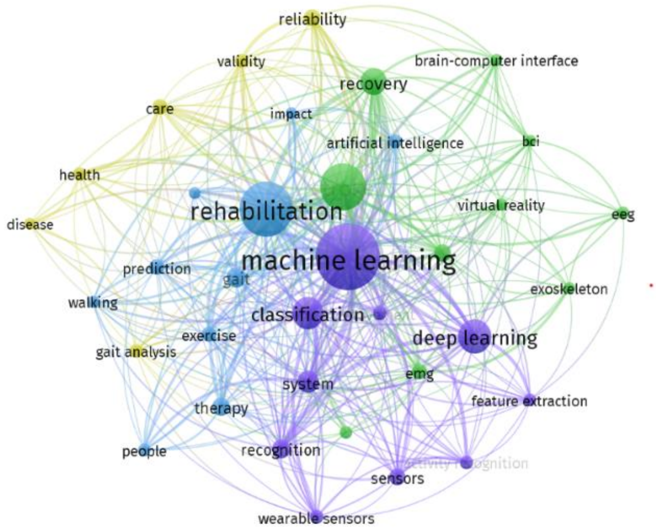


Fig. 16**.** Co-occurrence network of keywords

VosViewer uses network diagrams to illustrate the connections and highlight the strength and frequency of these collaborations. Each node of the network defines a certain entity like a keyword, author, journal, or country. Here the nodes represent keywords and the size of the node defines the frequency of occurrence.

Table 4. Occurrences and total link strength of the keywords

Keyword	Occurrences	Overall link strength
Machine learning	120	247
Rehabilitation	92	199
Stroke	74	186
Recovery	34	107
Classification	44	93
Deep learning	47	85
System	26	75
Recognition	21	59
Gait	23	54
therapy	20	51

Moreover, the co-occurrence of keywords is denoted by the link between nodes. A thicker line signifies a strong connection between keywords, while thinner lines indicate a weaker association. Keywords with no connecting lines show that no relationship has been established. So, from Table 4, we can see the top 10 link strengths of the keywords in accordance with their occurrence. Various colours divide them into various clusters. We have 4 thematic clusters of red, green, blue, and yellow which are named after clusters 1, 2, 3, and 4. We have 10 items in the first 3 clusters and 4 items in cluster 4. So, we can use the nodes as well as links to describe the relationship among different clusters[15].

3.3.1 *Author's collaboration*

Co-authorship indicates some questions: Why do authors collaborate from different disciplines? Do more productive authors collaborate or not? To analyze these answers, the author's collaboration network is introduced to represent the collaboration techniques of the authors[8]. Collaboration networks of the scholars link up the relationship among the scholars in the publication field. The collaboration index of our data is 4.86 and authors per document is 4.74.

Table 5. Collaboration of authors

Documents of Single-author	13
Publications per Author	0.211
Authors per Publication	4.74
Co-Authors per Publication	5.4
Collaboration Index	4.86

To visualize the cooperation of the authors with research close contacts, a collaboration network is formed in Fig. 17. Vosviewer software can generate a collaboration map of scholars of a certain field to find out the scientific cooperation between them. Vosviewer categorizes the authors in two different clusters with red and green colours under the rehabilitation field. The red one is cluster 1 which contains 16 nodes whereas the green one contains 2 nodes. Here's a total of 18 items in the two clusters where the total links are 137. The total link strength is 273 in the collaboration network. Here the author is taken as a unit of analysis and 25 authors per document is taken as the maximum number. The minimum threshold of documents of an author is taken as 2 documents. One of the green clustered nodes is only connected with the same cluster and the other node is connected with the red cluster. So, the total number of nodes is 158 but all of them are not connected internally. Accordingly, 18 items are chosen from the 158 items that are collaborating and we get the collaboration network. The circular nodes represent different scholars and the links between the nodes reflect the relation or collaboration between authors. Link strength between two nodes demonstrates the relation between them and the thickness of each link is proportional to the link strength. The red cluster is mostly visible and has strong links between themselves and cluster 1. It indicates that the scholars of the likely fields collaborated with themselves to form co-authorship.

On the other hand, the green cluster is also mostly linked up between the same clusters but also connected with the other one. Clicking on each node, we can see its collaboration network with the multiple links between other nodes. Similarly, link strength can be visualized by clicking on the links[16].

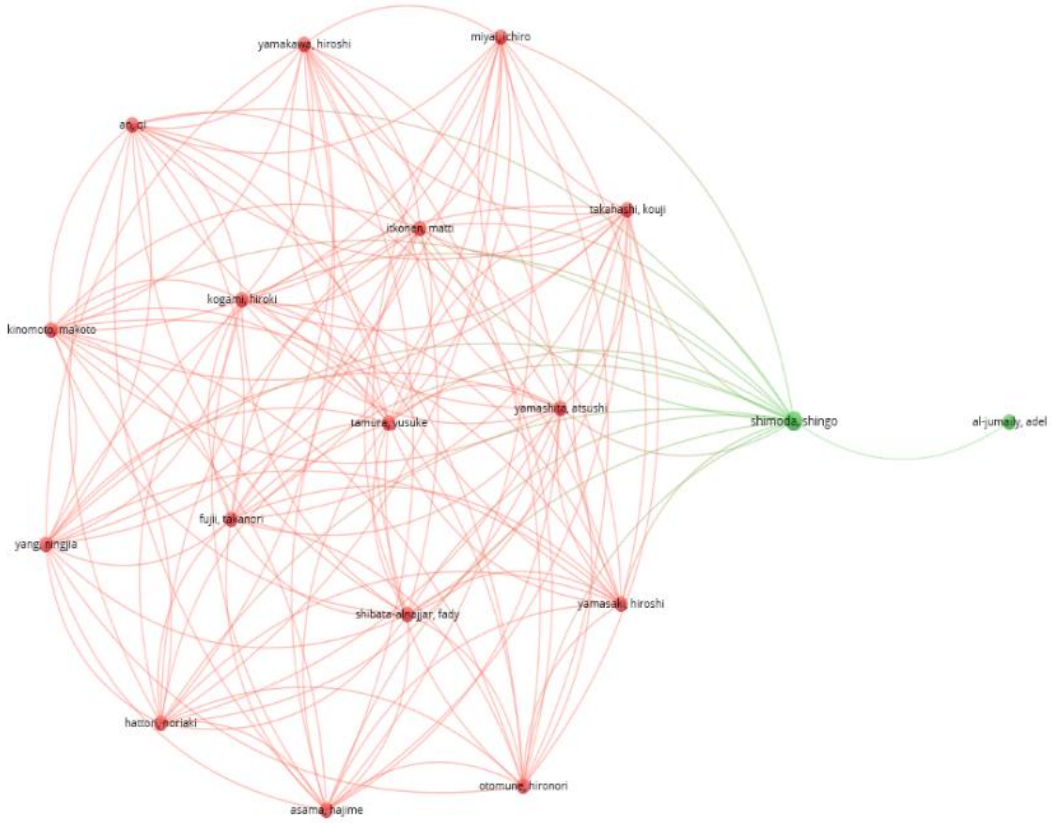


Fig. 17. Author's collaboration network

3.3.2 Country collaboration

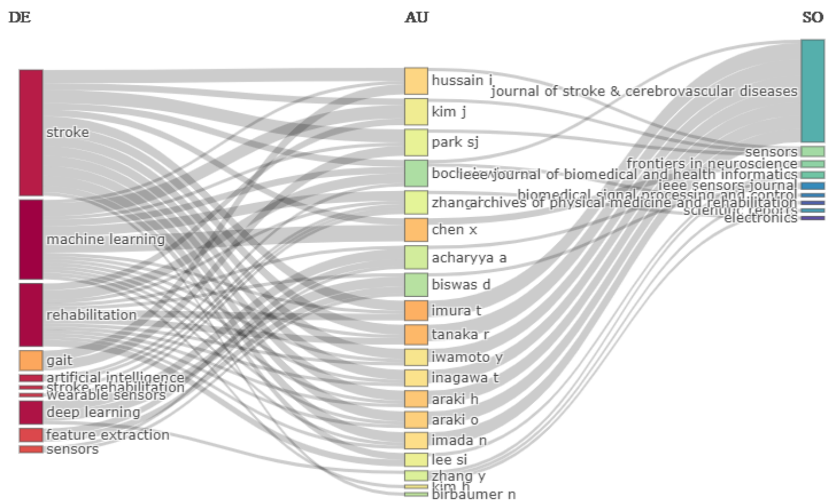
Authors belong to various countries and so the relationship between the author's countries builds up. The linkup between different countries is called country collaboration. Closeness and betweenness are two parameters to represent country collaborations. Here in the hierarchical arrangement, countries are according to the increasing values of the betweenness. Centrality measures for publications on board diversity and degree of centrality refer to the number of relational ties a document or author has in a certain network. Betweenness centrality represents a node's capability to carry information between other unconnected node groups inside a network, where each of the nodes symbolizes a certain document or author. In Fig. 18, the countries are divided into 4 clusters which are represented in 4 different colors. As clustering enriches bibliometric analysis by social or thematic clustering, several methods can be used to divide them into clusters such as exploratory factor analysis, hierarchical clustering, Island algorithm, Louvain method, and so on. Straight lines between the country circles indicate the relation between the countries[17].

Fig. 18. Country collaboration network (betweenness in ascending order)

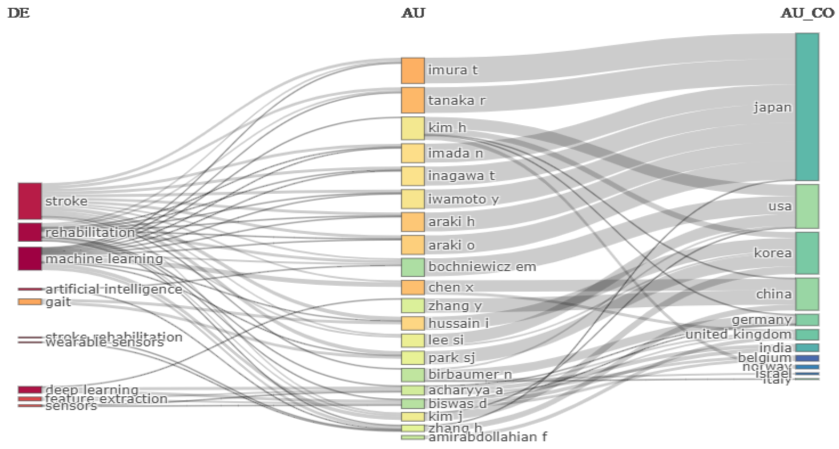
Table 6. Clusters and their closeness according to increasing betweenness

Betweenness	Node	Cluster	Closeness
0	Canada	1	0.021277
0	Japan	1	0.019608
0	U Arab emirates	1	0.019608
0	Italy	2	0.02381
0	Spain	2	0.018868
0	Israel	2	0.018519
0	Thailand	2	0.018868
0	Malaysia	3	0.017857
0	Belgium	3	0.017857
0.872549	Saudi Arabia	4	0.022727
0.96875	Norway	4	0.022222
5.781087	Korea	1	0.02439
22.83172	Germany	2	0.026316
31	Australia	1	0.028571
31.28847	India	3	0.025
31.57695	United Kingdom	2	0.027027
61.41844	China	1	0.033333
63.26203	USA	1	0.032258

From this geographic map, we can find out the most contributed and the less contributed countries as well.



(a)



(b)

Fig. 19(a). Shanky diagram relation between keywords, authors and sources, Fig. 19(b). Shanky diagram of keywords, authors and authors country

Clear visualization of the interrelation among the authors, keywords, source, and author's countries is demonstrated here. In the field of rehabilitation, this three-field plot reflects the correlation between the three parameters. In Fig. 19(a) of the three-field plot, the leftmost column represents the author's keyword, the middle one indicates the authors and the rightmost indicates sources of this rehabilitation field. This figure demonstrates how the authors and their keywords are connected and the strength of the connection. In the case of keywords, stroke is used by most of the authors than the other keywords in the figure. The height of the boxes ensures the importance or usefulness of the particular keyword or author. Machine learning is the second most dominating keyword and the following are rehabilitation, deep learning, gait, and so on. Representation of the sources or journals can be noticed from the rightmost column. Among the influential journals, the Journal of Stroke and Cerebrovascular Diseases has most of the affiliations by the authors. Sensors, The IEEE Sensor Journal follows the first one as they have a smaller number of author affiliations[7]. The second three fields plotted in 18(b) correlate the author's country with the authors as well as the keywords. Japan achieved a large number of affiliations among the other countries in this rehabilitation field. USA, Korea, and China gradually gained affiliations with the authors in this field. Among the author's keywords, stroke became the most dominant keyword, and rehabilitation, as well as machine learning, got the next higher volume of authors.

3.4 Trend analysis

This study summarizes the various trends followed by authors, including their countries of origin along a specific time period and the impact of them on the publications.

The search trends are illustrated in Fig. 20 in accordance with the analysed keywords. Here we have a trend map of 30 keywords for 10 years from 2011 to 2021. If we observe the map, we can see that plasticity and inpatient rehabilitation are the trending keywords in 2015.

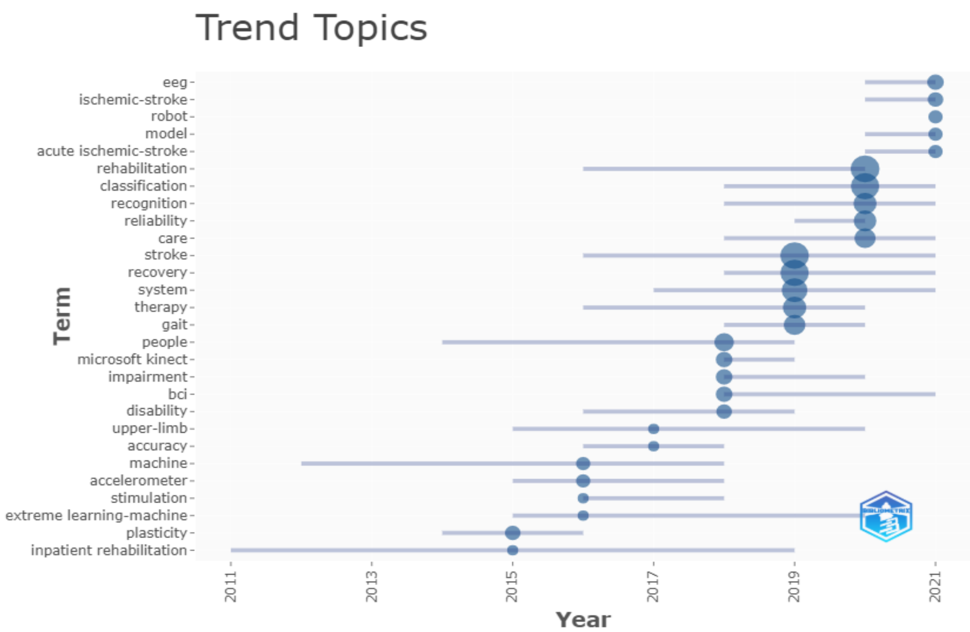


Fig. 20. Trending topics and thematic analysis of the field of rehabilitation

The following circles demonstrate the trending keywords and larger circles represent a higher frequency of occurrence. If we observe 2019 and 2020, we may have the largest circles there. Stroke and recovery became hot topics in 2019 whereas systems and therapy were also trending then. In 2020, rehabilitation, classification, recognition, reliability, and care were on the top but rehabilitation and classification had higher frequency. If we observe the trending keywords of 2021, EEG, ischemic-stroke, robot, model, and acute ischemic-stroke have become mostly searched topics. The horizontal lines describe the timescale of the topic being on trend.

3.4.1 Author based

This section focuses on authors as well as their influences on the collected articles for rehabilitation research. Author analysis mainly works to find out the significance of features related to authors. A total of 2026 authors are involved in this rehabilitation research work among which only 13 are sole-authored. The rest of the 2013 documents are multi-authored documents. The most relevant author representation was a part of the citation analysis that finds the most qualified and skilled authors in this rehabilitation field. Author qualification depends on the number of documents published by an author per year as well as the number of citations he receives per year. One author may get fewer citations while publishing more specialized documents in a certain field than another author with fewer publications[7].

Imura T and Tanaka R have 6 documents which is the highest number of documents among all the authors. Chen x is following with 5 documents and the rest of the authors have 4 or less numb of works in this field.

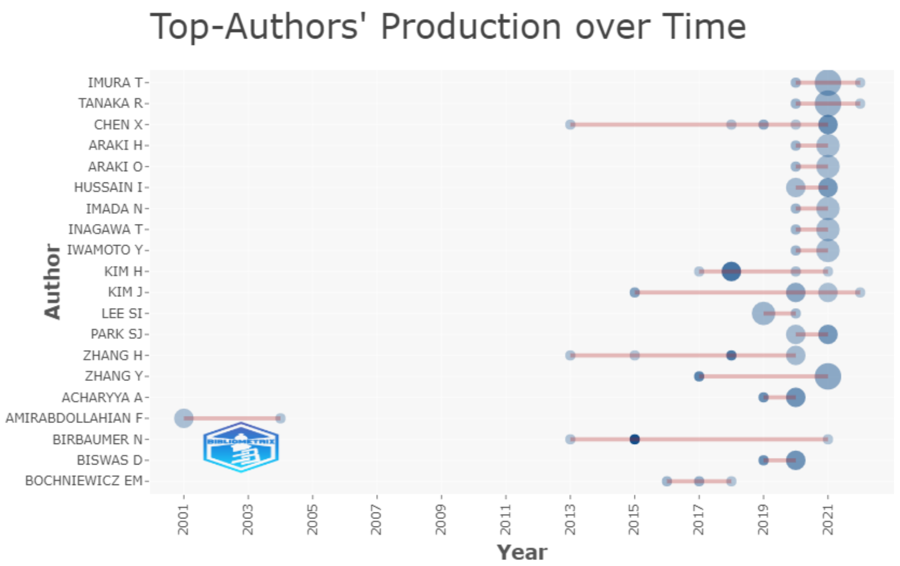


Fig. 21. Top author’s production over time

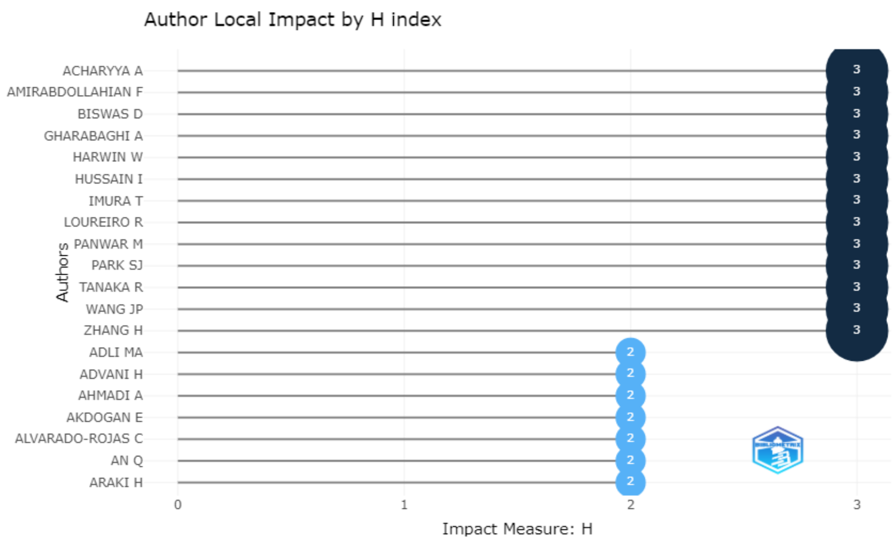


Fig. 22. Author local impact by H index

Productivity can also be estimated for individual authors in the field of rehabilitation. The top 20 author's productivity over the years 2001 to 2021 is demonstrated here in Fig. 21. The straight line denotes an author's working timeline. Bubble size indicates the number of publications produced yearly by a scholar. If the bubble is larger, it represents that the author has a large number of publications per year. On the other hand, the colour intensity of the bubble is proportional to the frequency of citations. Some of the authors started their production long ago and still working in this rehabilitation field whereas most of them started recently. Chen X, Zhang H, and Birbaumer N started earlier though their publications are a few. Imura T started his work in 2020 and continued for 2022. In 2021, he published the maximum number of papers though the number of average citations is less. Chen X is working from 2013 to 2021 with a smaller number of documents. He got a higher frequency of average citation in 2021 as the shade of the bubble is deeper. We can see that not many documents were published in the rehabilitation field before 2013. This figure visualizes that 2021 is the year when most of the documents are published and the citations are also high.

The h-index is a metric that reflects both the productivity and citation impact of a researcher's publications. It is determined by the number of the author's papers that have been cited at least that same number of times. The H-index function computes the h-index for authors or sources, along with its variants, such as the g-index and m-index, within a bibliographic dataset. Fig. 22 illustrates the h-index of the authors.

3.4.2 Country or region based

Publications according to regions or countries are described in this section with their single-country publication (SCP) and multiple-country publication (MCP) rates. China has 86 articles among which 68 SCP and 18 MCP.

Table 7. Scientific publication production by region/countries

Country	Articles	Freq	SCP	MCP	MCP_Ratio
CHINA	86	0.20283	68	18	0.2093
USA	81	0.19104	70	11	0.1358
ITALY	24	0.0566	17	7	0.2917
KOREA	22	0.05189	18	4	0.1818
UNITED KINGDOM	21	0.04953	16	5	0.2381
CANADA	20	0.04717	15	5	0.25
GERMANY	20	0.04717	17	3	0.15
INDIA	18	0.04245	10	8	0.4444
JAPAN	17	0.04009	16	1	0.0588
SPAIN	14	0.03302	11	3	0.2143
MALAYSIA	8	0.01887	4	4	0.5
AUSTRALIA	7	0.01651	5	2	0.2857
IRELAND	6	0.01415	4	2	0.3333
PORTUGAL	6	0.01415	4	2	0.3333
TURKEY	6	0.01415	6	0	0
COLOMBIA	5	0.01179	4	1	0.2
BRAZIL	4	0.00943	4	0	0
EGYPT	4	0.00943	2	2	0.5
ISRAEL	4	0.00943	2	2	0.5

Though the USA has fewer total publications than China, it has 70 single-country publications which are more than China. Italy, Korea, and the UK follow with 24, 22, and 21 articles whereas Brazil, Egypt, and Israel have the least publications.

3.4.3 Source or affiliation-based analysis

The study investigated the output of the publications of the most influential 20 affiliations in the rehabilitation field. UNIVERSITY WISCONSIN published 25 articles which is the maximum number of articles published among all the affiliations. NORTHWESTERN UNIVERSITY and UNIVERSITY TUBINGEN have 16 articles and the others are following. Among the top 20 affiliations, BOSTON UNIVERSITY, CATHOLIC UNIVERSITY AMER, and CHANG GUNG UNIVERSITY have the least number of articles.

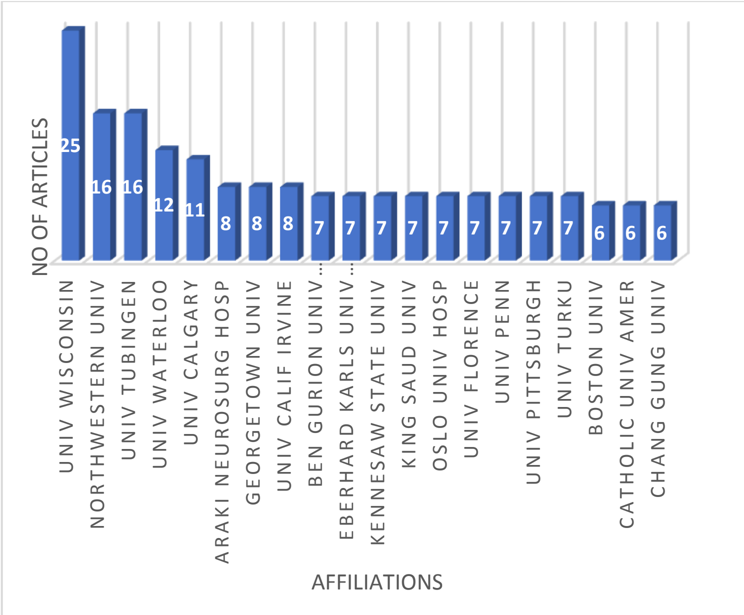


Fig. 23. Most relevant affiliations

Fig. 24 shows the top 20 relevant sources among the total 275 sources of this work. These top 20 sources are chosen according to their number of published articles in the rehabilitation field. This result is analyzed based on the data retrieved from Web of Science in March 2020. From the figure, we can get an idea regarding the sources and their impacts on this particular field of rehabilitation.

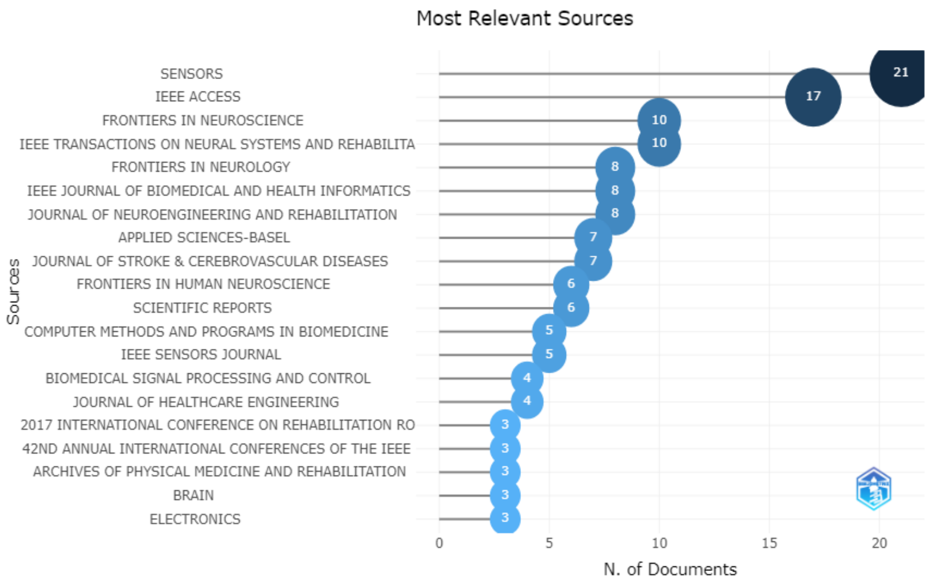


Fig. 24. most relevant sources

Sources are on the left side of the figure and their published number of documents is on the right side. "Sensors" is the most impactful journal among all of them where 21 documents are published there. "IEEE Access", and "Frontiers in Neuroscience" published 17 and 10 documents and the other sources have less than 10 published documents. Further investigation shows that "Brain" and "Electronics" have the least number of published documents among the top 20 prolific sources.

4. Conclusion

This paper provides a bird's eye view of the bibliometric analysis of the rehabilitation field with machine learning or deep learning. This study covers an extensive review of the rehabilitation field with scientific publications and analysis. This paper emphasizes establishing a proper knowledge of bibliometric analysis of the rehabilitation field among emerging scholars. Bibliometric analysis and visualization of the interrelations among the scholars, countries and sources contribute much to understand the correlations between them. The bibliometric analysis of the rehabilitation field is gaining much popularity as they offer to acquire vast knowledge regarding the particular field within a short time. However, due to the limited access of the softwires, we only used Biblioshiny and Vosviewer in our work which is one of the limitations of our work. Though many renowned scholars have worked in this rehabilitation field, bibliometric analysis of this field is not very popular. In this paper, individual author, country, and affiliation-based analysis is done so that scholars can easily get an overview from this bibliometric analysis. This analysis also depicted working trends of machine learning or deep learning-related research. Though there are currently a limited number of bibliometric analysis done in this field, future scholars will be dedicated to filling the gap in the rehabilitation field and this paper will help them to further analysis of this field.

References

[1] T. Tahsin, K. M. Mumenin, H. Akter, J. J. Tiang, and A.-A. Nahid, "Machine Learning-Based Stroke Patient Rehabilitation Stage Classification Using Kinect Data," *Applied Sciences*, vol. 14, no. 15, 2024, doi: 10.3390/app14156700.

[2] Y. Guo, Z. Hao, S. Zhao, J. Gong, and F. Yang, "Artificial intelligence in health care: bibliometric analysis," *J Med Internet Res*, vol. 22, no. 7, p. e18228, 2020.

[3] P. Srisusilawati, A. S. Rusydiana, Y. D. Sanrego, and N. Tubastuvi, "Biblioshiny R application on islamic microfinance research," *Library Philosophy and Practice*, vol. 2021, no. 5096, pp. 1–24, 2021.

[4] X. Xue *et al.*, "Global Trends and Hotspots in Research on Rehabilitation Robots: A Bibliometric Analysis From 2010 to 2020," *Front Public Health*, vol. 9, 2021.

[5] X. Feng, C. Liu, Q. Guo, Y. Bai, Y. Ren, and B. Ren, "Research progress in rehabilitation treatment of stroke patients: a bibliometric analysis. Neural Regen Res. 2013; 8 (15): 1423–30."

[6] J.-H. Huang, X.-Y. Duan, F.-F. He, G.-J. Wang, and X.-Y. Hu, "A historical review and Bibliometric analysis of research on Weak measurement research over the past decades based on Biblioshiny," *arXiv preprint arXiv:2108.11375*, 2021.

[7] F. J. Agbo, S. S. Oyelere, J. Suhonen, and M. Tukiainen, "Scientific production and thematic breakthroughs in smart learning environments: a bibliometric analysis," *Smart Learning Environments*, vol. 8, no. 1, pp. 1–25, 2021.

[8] M. Kabil, S. Priatmoko, R. Magda, and L. D. Dávid, "Blue economy and coastal tourism: A comprehensive visualization bibliometric analysis," *Sustainability*, vol. 13, no. 7, p. 3650, 2021.

[9] T. Tahsin *et al.*, "GWO-XGB: Grey Wolf Optimization-based eXtreme Gradient Boosting for Hypertension Prediction in Bangladesh," in *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)*, IEEE, 2021, pp. 1–4.

[10] Z. Xiao, Y. Qin, Z. Xu, J. Antucheviciene, and E. K. Zavadskas, "The Journal Buildings: A Bibliometric Analysis (2011–2021)," *Buildings*, vol. 12, no. 1, p. 37, 2022.

[11] A. S. Rusydiana, "Bibliometric analysis of journals, authors, and topics related to COVID-19 and Islamic finance listed in the Dimensions database by Biblioshiny," *Science Editing*, vol. 8, no. 1, pp. 72–78, 2021.

[12] S. Secinaro, V. Brescia, D. Calandra, and P. Biancone, "Employing bibliometric analysis to identify suitable business models for electric cars," *J Clean Prod*, vol. 264, p. 121503, 2020.

[13] W. Yan, X. Liu, B. Shan, and Y. Pu, "Research on the Emotions Based on Brain-Computer Technology: A Bibliometric Analysis and Research Agenda," *Front Psychol*, p. 4995, 2021.

[14] S. Fosso Wamba and M. M. Queiroz, "Responsible artificial intelligence as a secret ingredient for digital health: Bibliometric analysis, insights, and research directions," *Information Systems Frontiers*, pp. 1–16, 2021.

[15] T. Palma Pagano *et al.*, "Machine learning models and facial regions videos for estimating heart rate: a review on Patents, Datasets and Literature," *arXiv e-prints*, p. arXiv-2202, 2022.

[16] Y. Li, Z. Xu, X. Wang, and X. Wang, "A bibliometric analysis on deep learning during 2007–2019," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 12, pp. 2807–2826, 2020.

[17] J. Dores, T. Biju, and F. Sanu, "Mind the Gap: A Bibliometric Analysis on Unrecognized Role of Women in Fisheries Sector".

Authors' Profiles

Tasfia Tahsin was born in Jashore, Bangladesh, in 1998. She received the B.Sc. degree in Electronics and Communication Engineering from Khulna University of Bangladesh, in 2023. She is currently enrolling her M.Sc. from Khulna University. Her research interests include rehabilitation, machine learning, deep learning, artificial intelligence, big data analysis, machine fault classification.

Humayra Akter was born in Dhaka, Bangladesh, in 2000. She received the B.Sc. degree in Electronics and Communication Engineering from Khulna University of Bangladesh, in 2023. Her research interests include rehabilitation, machine learning, deep learning, artificial intelligence.

Uzzal Biswas received a Bachelor of Science degree in Electronics and Communication Engineering with distinction from Khulna University, Bangladesh in 2011. Later in 2011, he joined in Electronics and Communication Engineering (ECE) discipline as a faculty member. In 2014, he received a Master of Science degree in Electronics and Communication Engineering with distinction from Khulna University, Bangladesh. In 2015, he received a President Gold Medal Award from Khulna University for securing 1st position with distinction and obtaining the highest CGPA in the undergraduate program among all disciplines under the Science, Engineering and Technology (SET) School of Khulna University. In 2016, he received the prestigious Endeavour Postgraduate Research Scholarship from the Education Ministry of Australian Government to pursue his PhD degree at the University of New South Wales (UNSW), Sydney, Australia. In 2021, he received a Doctor of Philosophy degree in Biomedical Engineering from UNSW, Sydney, Australia. Currently, he is working as a Professor in the ECE discipline, Khulna University, Khulna, Bangladesh. His research interests are Biomedical Signal Processing, Telehealth-based Remote Patient Monitoring, Data Mining, Time Series Analysis, Machine Learning-based Classification and Prediction Modeling.

Jun Jiat Tiang received the degree in electronics engineering from Multimedia University, Malaysia, the master's degree from the University of Science, Malaysia, and the Ph.D. degree from Universiti Kebangsaan Malaysia (UKM). He worked as a Design Automation Engineer with the Chipset Structural Design Team, Intel Microelectronics (M) Sdn. Bhd., Penang, Malaysia, from September 2004 to July 2005, and an Electronics Engineer with the Global Technology Development Group, Motorola Technology Sdn. Bhd., Penang, from May 2006 to May 2007. He has vast experience while working as the Project Leader in various research grants, such as TM Research and Development, from 2017 to 2020; Research Grant, Ministry of Science, Technology and Innovation (MOSTI), from 2008 to 2010; and research grant Mini Fund, in 2016. He is currently a Senior Lecturer and a Researcher with the Faculty of Engineering, Multimedia University. His research interests include RFID, microwave circuits, antenna, and propagation. He was awarded the Gold Medal at the 23rd International Invention, Innovation, and Technology Exhibition (ITEX) 2012, Kuala Lumpur, Malaysia, in May 2012, and the Silver Medal at the Malaysian Technology Expo (MTE) 2013, Kuala Lumpur, in February 2013.

Abdullah-Al Nahid received the B.Sc. degree in Electronics and Communication Engineering from Khulna University, Khulna, Bangladesh, in 2007, the M.Sc. degree in telecommunication engineering from the Institute for Telecommunication Research, University of South Australia, Mawson Lakes, SA, Australia, in 2012, and the Ph.D. degree in machine learning from Macquarie University, Sydney, NSW, Australia, in 2018. He is currently with the Electronics and Communication Engineering Discipline, Khulna University. His research interests include machine learning, biomedical image processing, data classification, and smart grid.

How to cite this paper: Tasfia Tahsin, Humayra Akter, Uzzal Biswas, Jun Jiat Tiang, Abdullah-Al Nahid, "A Comprehensive Bibliometric Study on Machine Learning Based Rehabilitation and Stroke Research (1999 - 2022)", International Journal of Engineering and Manufacturing (IJEM), Vol.15, No.1, pp. 11-31, 2025. DOI:10.5815/ijem.2025.01.02