Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.9854933

BIG DATA ANALYTICS & HEALTHCARE- LITERATURE REVIEW & BIBLIOMETRICS

Aditi Vivek Mishal
Chief Sustainability Officer & Dean, Govardhan Ecovillage

Abstract
Big Data Analytics has emerged as a powerful tool in healthcare, enabling the analysis of large-scale data to uncover valuable insights for patient care, disease prevention, and resource allocation. This paper explores the potential of Big Data Analytics in revolutionizing the healthcare industry by integrating electronic health records, wearable devices, and medical imaging to facilitate personalized medicine and evidence-based decision making. However, ethical considerations, privacy concerns, and data security issues must be addressed to ensure responsible data usage. Collaboration between data scientists, healthcare professionals, and policymakers is crucial to leverage the full potential of Big Data Analytics and realize its transformative impact on patient outcomes and the future of healthcare. The paper aims at presenting exhaustive literature review of 141 documents derived from Web of Science database, Scopus and Google Scholar. Paper uses Bibliometrics which is a quantitative analysis method that focuses on evaluating and measuring the impact, visibility, and scholarly influence of scientific publications and researchers through various indicators such as citation counts, h-index, and journal impact factors. This paper concludes by highlighting the need for ongoing research and advancements in the field to maximize the benefits of Big Data Analytics in healthcare while safeguarding patient privacy and data integrity.

Introduction:

20th Century has witnessed exponential growth in Healthcare, and 21st century is argued to be evidence to Humans and machines working together for better health, with combination of human reasoning and deep learning (Raghupati W, Raghupati V, 2014; Belle A., Thiagarajan R, Soroushmehr, S, Naividi, Najarian, 2015; Morgan, 2017). With popularity of big data applications and particularly mhealth in Rural health being on consistent rise, authors felt it was necessary and relevant to give a detailed bibliometric analysis of publications. This section provides the analytical overview from different perspectives of bibliometric analyses. Literature review section utilizes a more comprehensive hybrid approach by providing concept centric descriptive review supported by visualization empowered bibliometric analyses. Thus, it will be equally enlightening learning experience for Researchers to have visually powered keyword co-occurrences, author co-citation analysis and network, productive authors, countries and their collaborative clusters. This book chapter will definitely add great value with background that very few literature review articles exist and lack the bibliometrics focus (Google scholar, 2018).

Web of Science database, Scopus and Google Scholar were used as database to carry out scientific enquiry. The syntax for search purposes was mhealth+Rural healthcare. The initial searches resulted into total of 381 articles in Scopus and 141 articles in Web of Science. The search was further narrowed for relevance with following sequential iterations.

1. Publications in English

2. Articles from Journal, conference proceedings and book review or chapters.
3. Articles published 2010 onwards, so this criteria was untouched
4. Articles from relevant fields were sorted like Medical, Nursing, Medicines, IT, Computational science etc.

Scopus and Web of Science search results were compared for degree of relevance. Post detailed study; it was decided to use Web of Science search results for detailed bibliometric analysis. However, other databases have supported for a comprehensive descriptive analysis for literature review.

Bibliometric as Research Methodology for Literature Review

Bibliometrics, as a research method, offers a quantitative approach to analyzing and assessing scholarly publications and their impact within the academic community. By employing bibliometric techniques, researchers can uncover patterns, trends, and relationships in scientific literature, facilitating evidence-based decision making and knowledge discovery. One commonly used bibliometric indicator is citation analysis, which examines the number of times a publication has been cited by other works, serving as a measure of its influence and significance (Garfield, 1979). Other indicators include the h-index, which combines the number of publications and their respective citation counts to assess an author's productivity and impact (Hirsch, 2005), and journal impact factors, which gauge the average number of citations received by articles published in a particular journal (Garfield, 2006). Bibliometrics provides researchers with a systematic and quantitative approach to evaluate research output, track scientific trends, and inform decision making in various fields (Waltman, 2016). However, it is important to note that bibliometrics should be used alongside qualitative methods to obtain a comprehensive understanding of the research landscape (Rafols et al., 2012). By combining bibliometric analysis with other research methodologies, researchers can gain valuable insights into the impact, visibility, and scholarly influence of scientific publications, contributing to evidence-based research evaluation and strategic planning.

Bibliometric analysis using Bibliometrix of R Studio:

The bibtex file was analysed using ‘Bibliometrix’ package as part of R Studio software and R software. Authors found Bibliometrix as very strong, scientifically grounded tool for effective bibliometric analysis. After initial installation of all related softwares; file path was written for access of bibtex file which was output of web of science search with above mentioned steps (R Core Team, 2018).

Records for 141 documents post above mentioned filtration served as feed for bibliometric analysis from sources as Journals, books, and conference proceedings totaling to 65 in numbers. The effective period selected was 2010-2018. Total number of varied keywords used by authors in selected research documents being 455. Average citations per document amount to 10.7. Collaboration index as calculated by programme in R Studio is 5.88; with 5.78 authors on an average writing one document. More details can be referred to in following table.

Table

Documents	141
Sources (Journals, Books, etc.)	65
Keywords Plus (ID)	397

Author's Keywords (DE)	455
Period	2010 - 2018
Average citations per documents	10.7
Authors	815
Author Appearances	958
Authors of single-authored documents	4
Authors of multi-authored documents	811
Single-authored documents	3
Documents per Author	0.173
Authors per Document	5.78
Co-Authors per Documents	6.79
Collaboration Index	5.88

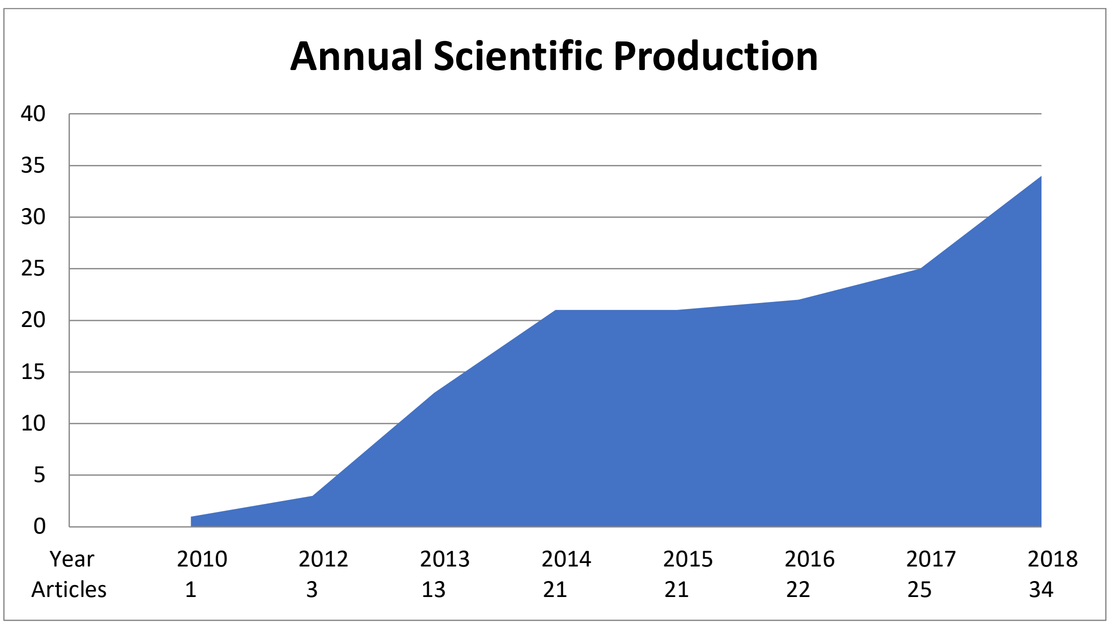

Fig. Number of Research documents published year wise for mHealth in Rural Healthcare

The annual scientific production saw a sudden jump in 2013, which can be greatly correlated with rise in applications on big data analytics particularly in healthcare, publication highest number appears in 2018 with approximately 34 publications and graph shows a rising trend.

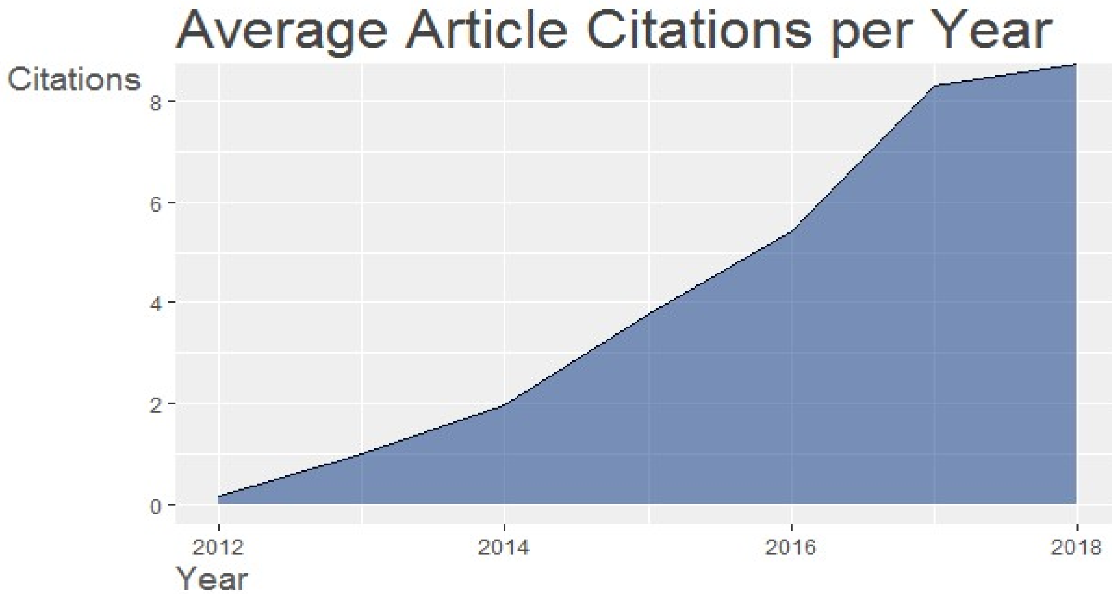

Fig. Average article citations per year for for mHealth in Rural Healthcare

With consistent rise in number of publications a consistent similar rising trend is observed in citations per document per year with highest being 9 citations in year 2018.

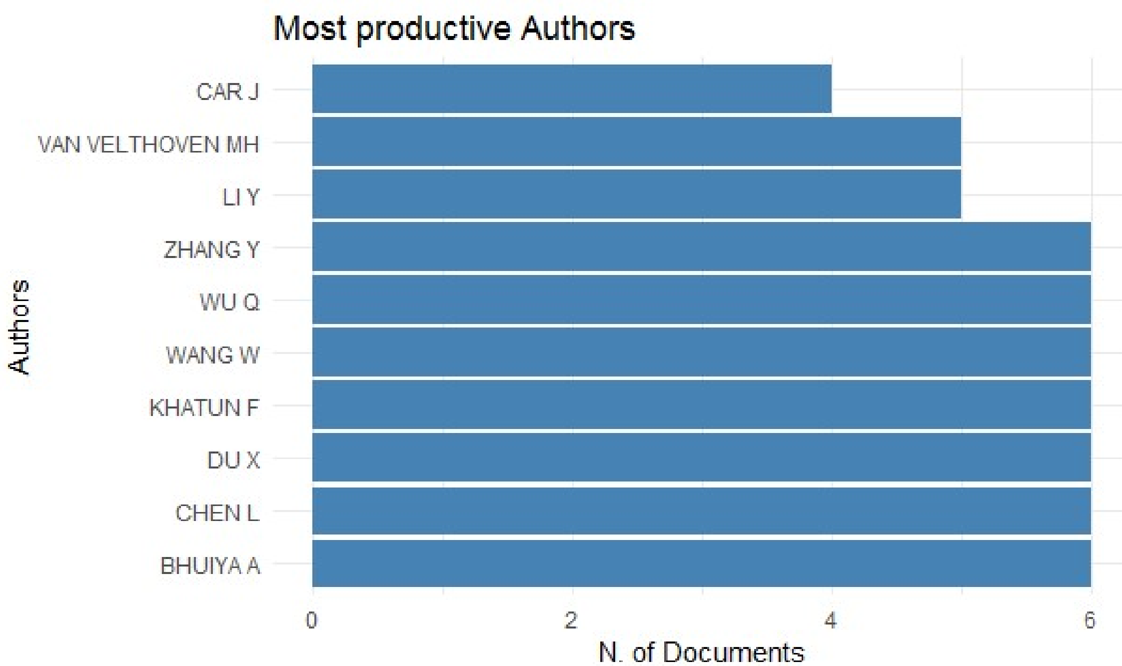

Fig. Topmost Authors for mHealth in Rural Healthcare

The scientific enquiry of topmost contributors in mHealth in Rural healthcare resulted in above graph, with top 7 contributors being Bhuiya, Chen, Du, Khatun, Wang, Wu, Zhang with 6 publications each. However, a deeper look into citations per Author changes this ranking drastically. Kaellander K (2013) has 166 citations for an article published in Journal of Medical and Internet Research followed by Aranda (2014) has 116 citations for an article published in BMC Public Health.

Table -Topmost Citations as per Contributors in Research Document

Paper	Total citations	Total citations/year
1 KAELLANDER K, 2013, J MED INTERNET RES	166	33.20
2 ARANDA-JAN CB, 2014, BMC PUBLIC HEALTH	116	29.00
3 WEINSTEIN RS, 2014, AM J MED	103	25.75
4 HALL CS, 2014, GLOB HEALTH ACTION	78	19.50
5 MAHMUD N, 2010, TECHNOL HEALTH CARE	75	9.38
6 LUND S, 2012, BJOG	74	12.33
7 WESOLOWSKI A, 2012, PLOS ONE	70	11.67
8 MARS M, 2013, PROG CARDIOVASC DIS	44	8.80
9 BLOOMFIELD GS, 2014, GLOBAL HEALTH	43	10.75
10 PRICE M, 2013, J MED INTERNET RES	35	7.00

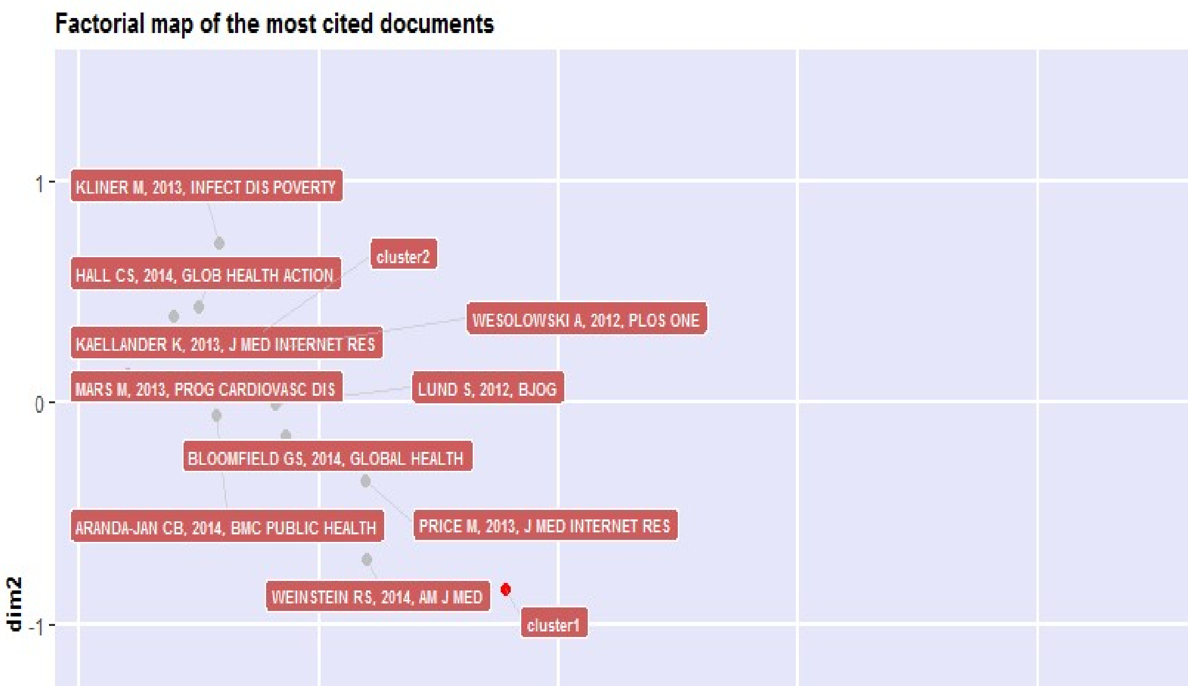

Fig. Factorial Map of most cited Documents for mHealth in Rural Healthcare

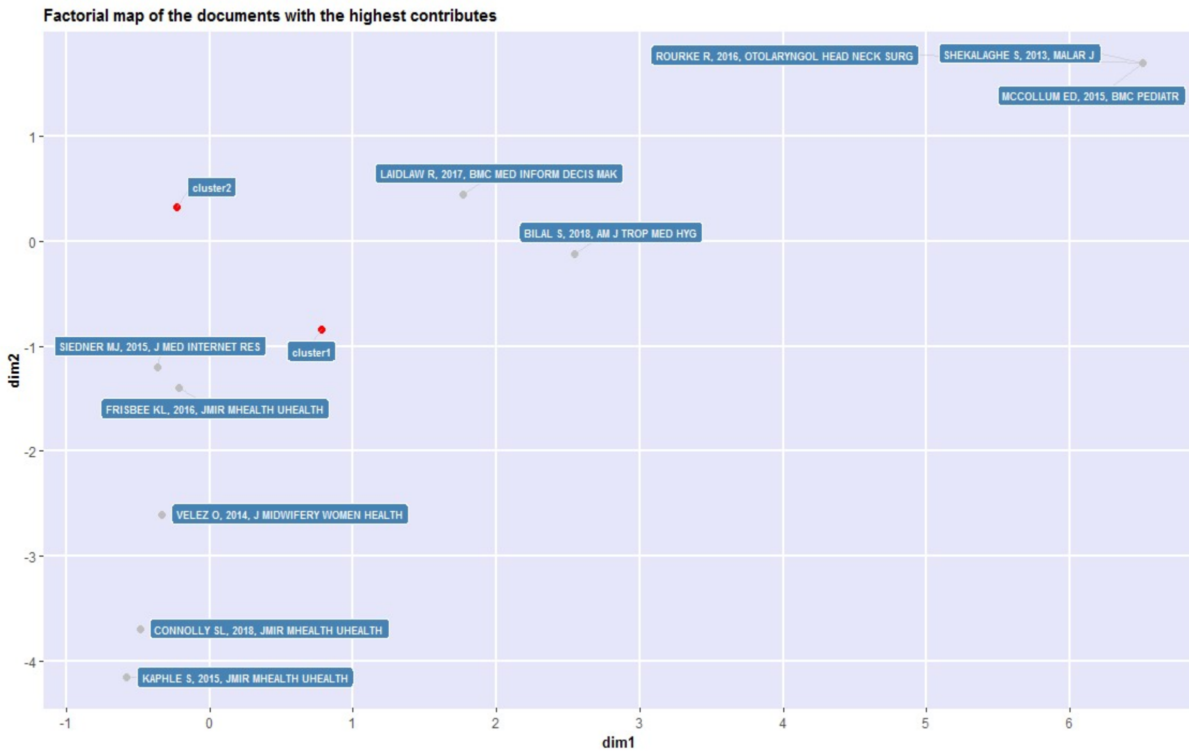

Fig. Factorial Map of the documents with highest contributes for mHealth in Rural Healthcare

There are four approaches popularly used in Citation analysis- co-citation analysis, bibliographic coupling, direct citation, and a bibliographic coupling-based citation-text hybrid approach. The last one is considered as more effective than other pure citation-based approaches even more than bibliographic coupling (Boyack K., Klavans R., 2010). Above 2 figures illustrate the factorial maps for highest cited and highest contributing documents across 2 clusters of co-cited articles which is elucidated in figure below. Moreover, the figure which follows below helps us to understand the effectiveness of bibliographic coupling citation-based approach as compared to direct citation network as depicted below.

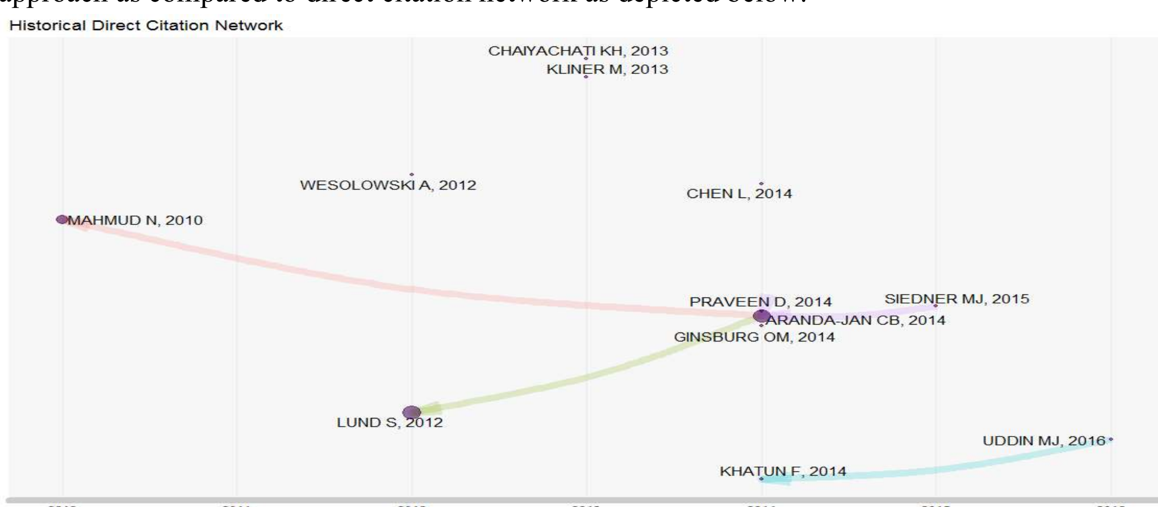

Fig. Direct Citation Network for mHealth in Rural Healthcare

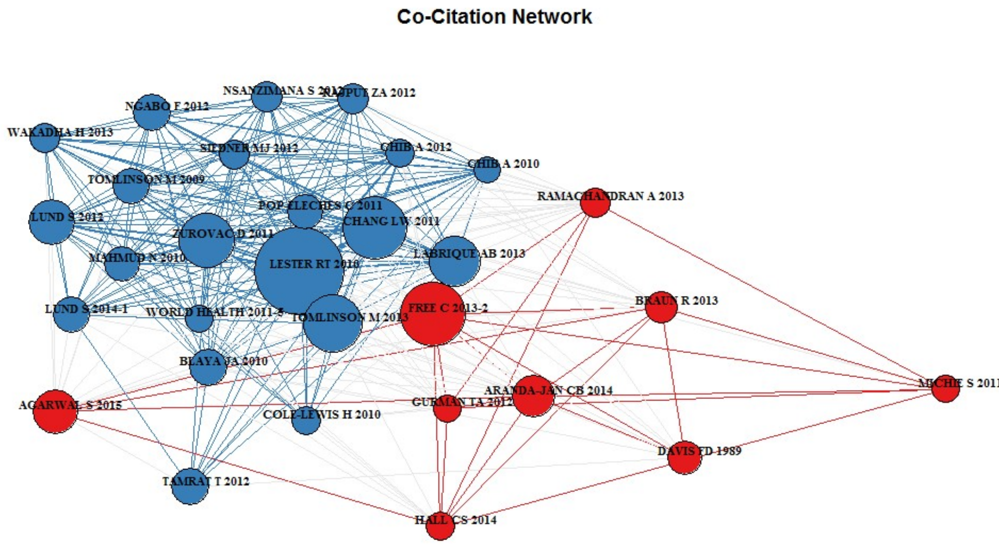

Fig. Co-citation Network for mHealth in Rural Healthcare

Fig. Detailed Co-citation network analysis using Gephi for mHealth in Rural Healthcare

The research authors as per documents to the centre of the network represent highly impactful articles with highest citations, the number of citations go on reducing as we move away from the centre towards the edges in the network. The co-citation detailed network very clearly charts out the 6 major clusters as mentioned earlier. As many connections from a particular document node those many co-citations connections, thus in a clutter of central network we can observe several for highly cited documents.

Table : Top Pageranked articles cited by other highly cited articles for mHealth and Rural Healthcare

Top Pageranked articles cited by other highly cited articles		
Rank	**Article**	**Pagerank**
1	Hall CS; Fottrell E; Wilkinson S; Byass P	0.018754
2	Aranda-Jan CB; Mohutsiwa-Dibe N; Loukanova S	0.018458
3	Prieto JT; Zuleta C; Rodriguez JT	0.018162
4	Lodhia V; Karanja S; Lees S; Bastawrous A	0.01786
5	Siedner MJ; Santorino D; Haberer JE; Bangsberg DR	0.0168
6	Hackett K; Lafleur C; Nyella P; Ginsburg O; Lou W; Sellen D	0.016312
7	van Velthoven MH; Li Y; Wang W; Du XZ; Wu Q; Chen L; Majeed A; Rudan I; Zhang YF; Car J	0.015528
8	Lund S; Boas IM; Bedesa T; Fekede W; Nielsen HS; Sorensen BL	0.015341
9	Ggita JM; Ojok C; Meyer AJ; Farr K; Shete PB; Ochom E; Turimumahoro P; Babirye D; Mark D; Dowdy D; Ackerman S; Armstrong-Hough M; Nalugwa T; Ayakaka I; Moore D; Haberer JE; Cattamanchi A; Katamba A; Davis JL	0.015249
10	van Velthoven MH; Li Y; Wang W; Chen L; Du XZ; Wu Q; Zhang YF; Rudan I; Car J	0.014678

Above table depicts an interesting list of articles. Gephi also allows to carry out Page ranking in co-citation network analysis. The page ranking gives interesting perspective of those articles which have been cited by highly cited articles. This analysis in Gephi resulted in a above top 10 articles. The top most being Hall et.al, followed by Aranda-Jan et.al., Prieto et.al., Lodhia et.al., Siedner et.al., Hackett et.al., etc.

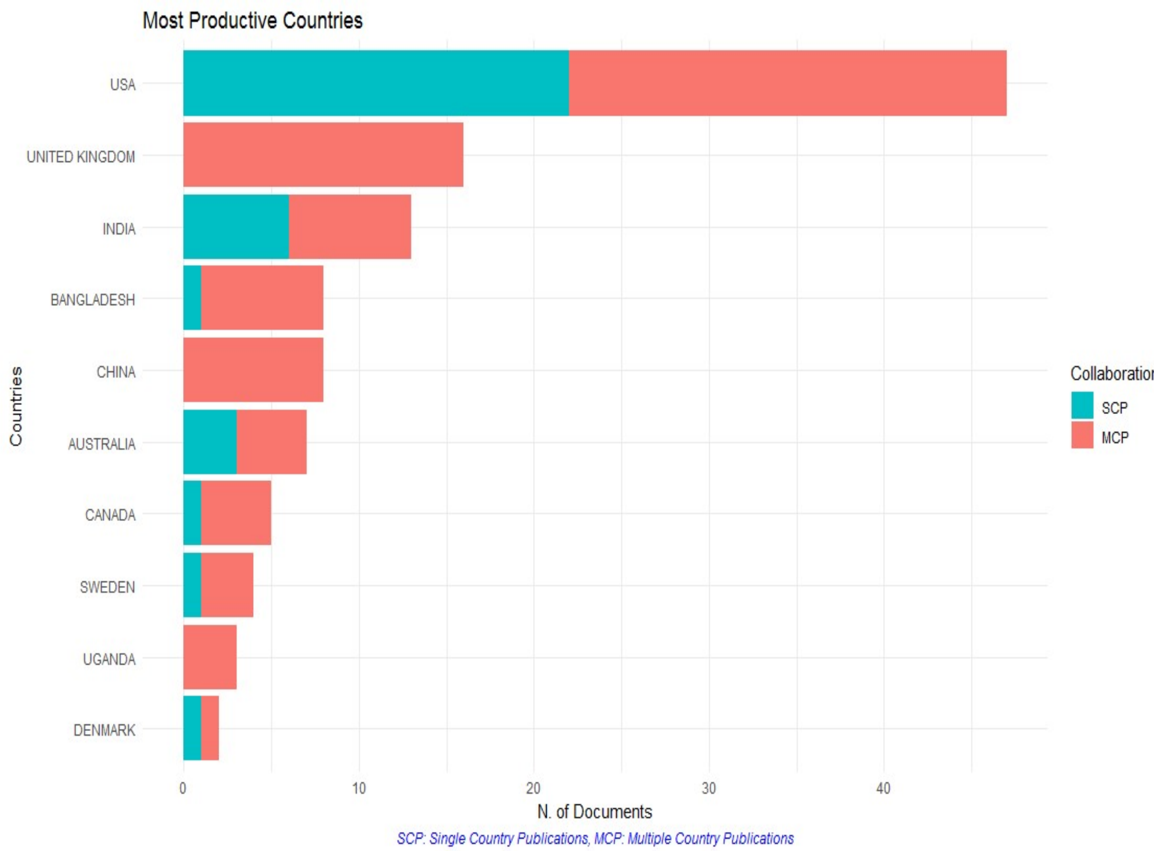

Fig. Country wise Publications for Big data Analytics in Healthcare

USA with over 50, followed by UK with 16 which is followed by India with 14 articles remain the 3 topmost contributing countries for publications in mHealth in Rural healthcare. Bangladesh and China follow with 8 publications. Graph also educates us on the across country collaborations for these publications with China showing 100% collaboration for publications. Following figure gives more details on collaborations across countries. A big cluster appears which is close knit in terms of collaboration which is USA, UK, Australia, India and Bangladesh and as was evident in graph above, the collaborations are stronger again with China, Kenya, Canada, Uganda, Ghana, and Spain which are other influential contributors.

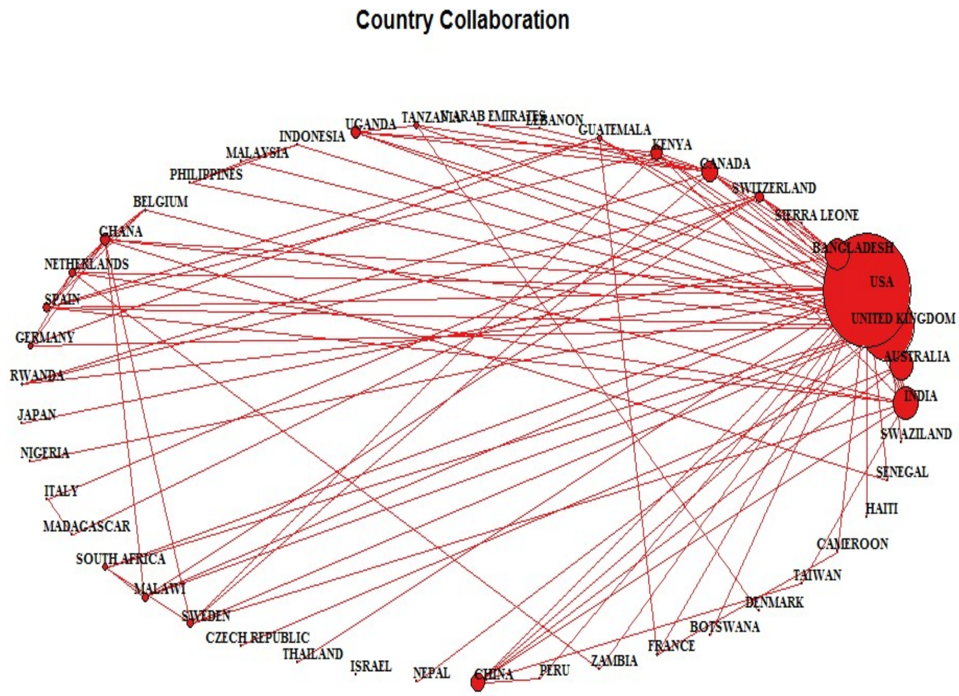

Fig. Country Collaborations for mHealth and Rural Healthcare

Table -Total Citations and average article citations per Country

Rank	Country	Total Citations	Avg Article Citations
1	USA	539	11.47
2	UGANDA	168	56.00
3	UNITED KINGDOM	167	10.44
4	GERMANY	116	58.00
5	DENMARK	88	44.00
6	BANGLADESH	82	10.25
7	INDIA	72	5.54
8	CANADA	56	11.20
9	AUSTRALIA	49	7.00
10	SOUTH AFRICA	44	22.00

Table above depicts citations per article as per country gives different ranking for top 3 contributors- USA with 539, Uganda with 168 and UK with 167 total citations. Moreover, the list of countries which has higher average citation per year includes Germany, Uganda and Denmark, with 58.00, 56.00 and 44.00.

Table -Topmost Sources for mHealth and Rural Healthcare

Rank	Sources	Articles
1	JMIR MHEALTH AND UHEALTH	15
2	BMC MEDICAL INFORMATICS AND DECISION MAKING	11
3	JOURNAL OF MEDICAL INTERNET RESEARCH	9
4	PLOS ONE	9
5	GLOBAL HEALTH ACTION	7
6	BMC PREGNANCY AND CHILDBIRTH	5
7	BMC PUBLIC HEALTH	5
8	INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS	5
9	BMC HEALTH SERVICES RESEARCH	3
10	BMJ OPEN	3

As depicted in above table, topmost contributing sources for research documents in mHealth and Rural healthcare are JMIR mHealth and uHealth with 15, BMC Medical Informatics and Decision Making with 11 and Journal of Medical Internet Research and PLOS One with 9 publications each.

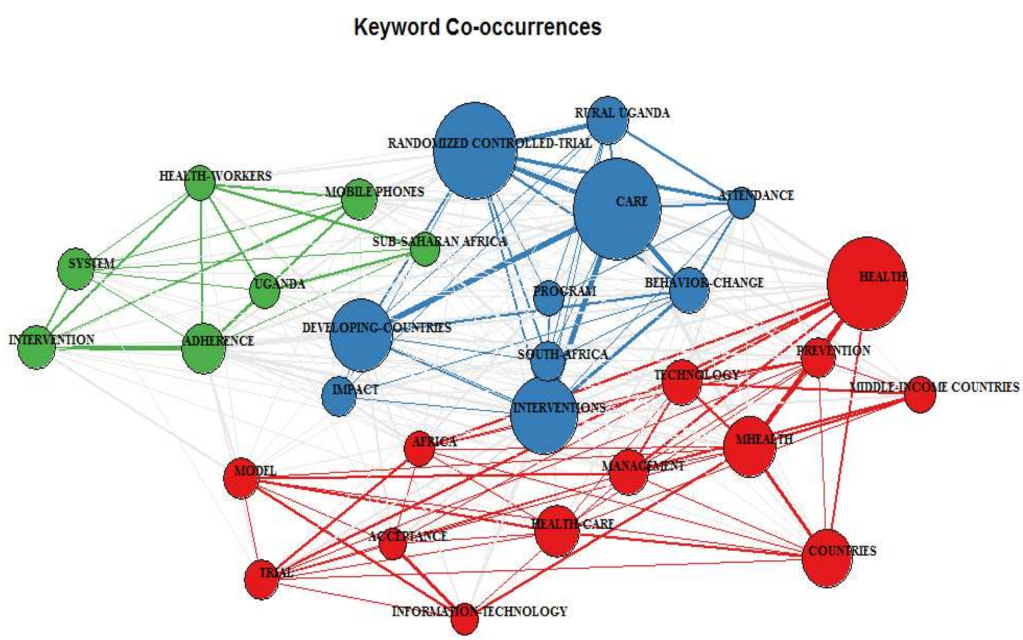

Fig. Keyword co-occurrences for mHealth and Rural Healthcare

In the list of 455 different words appearing as keywords in the research documents, keyword co-occurrences network is depicted in above figure. Words like mHealth, Telemdeicine, Care, Randomized controlled trail, Developing Countries, and interventions appear as prominent with 3 distinct co-occurrence clusters. The frequency of keyword occurrence is depicted by the size of circle which indicates the strength of that node in the network.

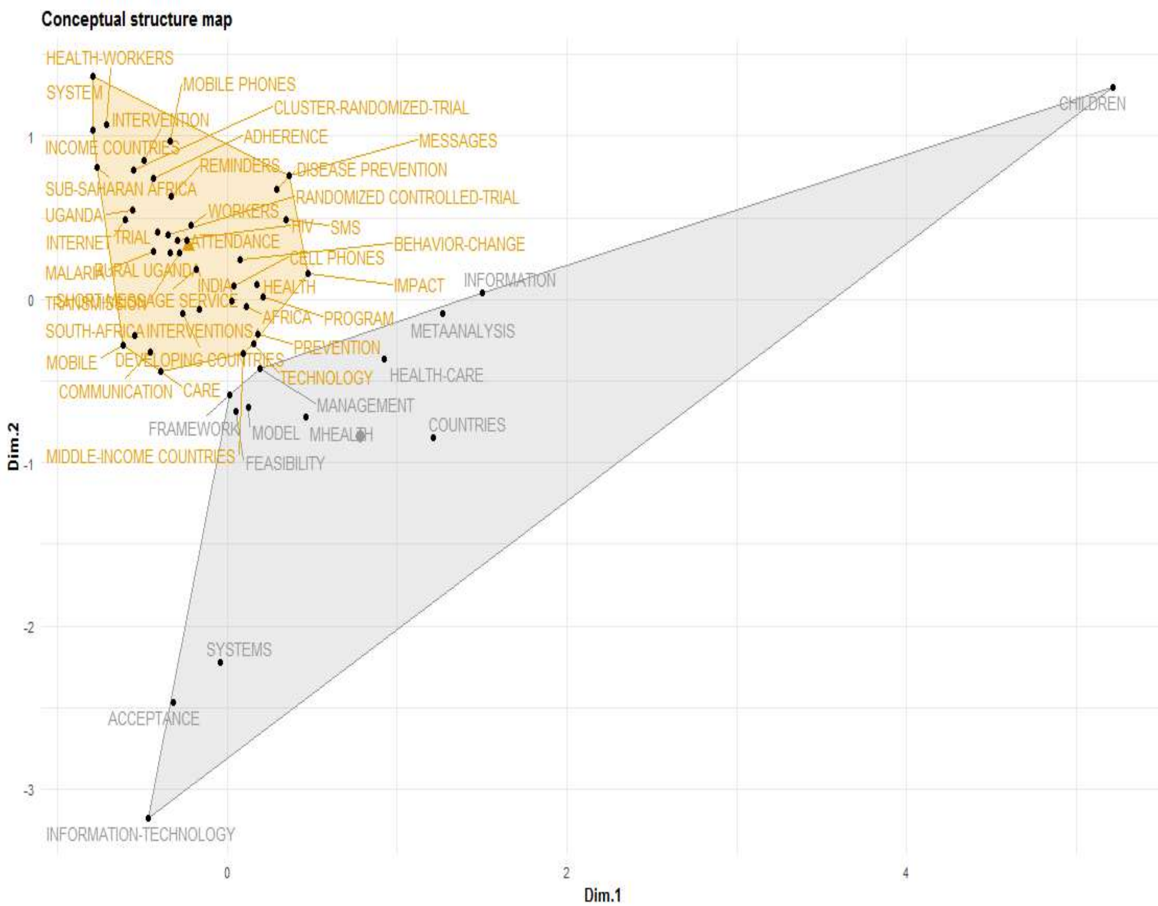

Fig. Conceptual Structure Map for mHealth and Rural Healthcare

Conclusion:

In conclusion, Big Data Analytics holds immense promise in revolutionizing the healthcare industry. By harnessing the power of large-scale data analysis, we can gain valuable insights into patient care, disease prevention, and resource allocation (Ozdemir and Saba, 2020; Sacchi et al., 2018). Through the integration of electronic health records, wearable devices, and medical imaging, we can unlock new opportunities for personalized medicine and evidence-based decision making (Xu et al., 2018; Choi et al., 2020).

However, it is important to address ethical considerations, privacy concerns, and data security issues to ensure the responsible use of Big Data Analytics in healthcare (Kayaalp, 2019; El Emam et al., 2021). As technology continues to advance, the collaboration between data scientists, healthcare professionals, and policymakers becomes crucial for leveraging the full potential of Big Data Analytics and realizing its transformative impact on improving patient outcomes and shaping the future of healthcare (Koh and Tan, 2017; Sittig et al., 2020).

References:

1. Agrawal, N., 2016. “Inequality in India: what’s the real story.” World Economic Forum, 4 October. Evaluation et al., 2016.
2. Aijaz, S., 2015. “6 social development efforts that will take India forward,” Your Story, 29 January.

3. AlphaBeta analysis for the Business and Sustainable Development Commission
4. Belle, A., Thiagarajan, R., Soroushmehr, S. M., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big data analytics in healthcare. *BioMed research international*, *2015*.
5. Boyack K., Klavans R., (2010) Co-Citation Analysis, Bibliographic Coupling, and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately? Journal of the American Society for Information Science and Technology 61(12):2389-2404 DOI: 10.1002/asi.21419
6. Choi, Y., Park, J., Lee, J., & Lee, S. (2020). Big Data Analytics for Smart Healthcare: Challenges and Solutions. Journal of Medical Systems, 44(10), 176. doi:10.1007/s10916-020-01690-9
7. El Emam, K., Jonker, E., Arbuckle, L., Malin, B., & Pomares, G. (2021). A Systematic Review of Re-Identification Attacks on Health Data. PLOS ONE, 16(1), e0247267. doi:10.1371/journal.pone.0247267
8. Garfield, E. (1979). Citation Indexing: Its Theory and Application in Science, Technology, and Humanities. Wiley.
9. Garfield, E. (2006). The History and Meaning of the Journal Impact Factor. JAMA, 295(1), 90–93. doi:10.1001/jama.295.1.90
10. Google Schoalr (2018) https://scholar.google.co.in/scholar?hl=en&as_sdt=0%2C5&q=literature+review+big+data+analytics+in+healthcare&btnG=, accessed Oct. 2018
11. Government of India, Ministry of Statistics and Programme Implementation. “Statistical Year Book India 2016.”
12. Hirsch, J. E. (2005). An Index to Quantify an Individual's Scientific Research Output. Proceedings of the National Academy of Sciences, 102(46), 16569–16572. doi:10.1073/pnas.0507655102
13. Kaka, N., Madgavkar, A., Manyika, J., Bughin, J., and Parameswaran, P., 2014. India’s technology opportunity: Transforming work, empowering people. McKinsey Global Institute.
14. Kayaalp, M. (2019). Big Data Analytics in Healthcare: Promise and Potential. Health Information Science and Systems, 7(1), 4. doi:10.1007/s13755-018-0061-4
15. Koh, H. C., & Tan, G. (2017). Data Analytics for Personalized Healthcare. Singapore Medical Journal, 58(9), 505-510. doi:10.11622/smedj.2017112
16. Morgan J. (2017) 7 ways big data analytics can boost healthcare, https://www.beckershospitalreview.com/healthcare-information-technology/7-ways-big-data-analytics-can-boost-healthcare.html, accessed Oct. 2018
17. Nambiar, R., Bhardwaj, R., Sethi, A., & Vargheese, R. (2013, October). A look at challenges and opportunities of big data analytics in healthcare. In *2013 IEEE international conference on Big Data* (pp. 17-22). IEEE.
18. Ozdemir, V., & Saba, V. K. (2020). Big Data Analytics for Genomic Medicine. In Big Data Analytics for Genomic Medicine (pp. 1-13). Academic Press. doi:10.1016/b978-0-12-817379-3.00001-6

19. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

20. Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How Journal Rankings Can Suppress Interdisciplinary Research: A Comparison between Innovation Studies and Business & Management. Research Policy, 41(7), 1262–1282. doi:10.1016/j.respol.2012.03.015

21. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, *2*, 1-10.
22. Sacchi, L., Gabaglio, S., de Gregorio, E., & Trucco, G. (2018). Big Data in Healthcare: Approaches, Challenges and Prospects. Yearbook of Medical Informatics, 27(01), 199-206. doi:10.1055/s-0038-1667084
23. Sittig, D. F., Murphy, D. R., & Smith, M. W. (2020). Identifying Areas for Improvement in Electronic Health Record-generated Injury Event Narratives in Big Data Analytics Environments. Journal of the American Medical Informatics Association, 27(4), 602-608. doi:10.1093/jamia/ocaa032
24. Surie, M.D., 2015. "Why India Could Make or Break the Success of SDGs." Asia Foundation, 14 October.
25. UNDP (2018) About India, http://www.in.undp.org/content/india/en/home/countryinfo.html, accessed Sept. 2018
26. Waltman, L. (2016). A Review of the Literature on Citation Impact Indicators. Journal of Informetrics, 10(2), 365–391. doi:10.1016/j.joi.2016.02.007
27. WHO, 2014. Health Expenditure, public % GDP.
28. World Bank, 2016. "Poverty and Shared Prosperity."
29. Xu, J., Wang, F., & Yuan, X. (2018). Clinical Big Data and Real-World Data Analytics for Precision Medicine. In Clinical Precision Medicine (pp. 61-83). Academic Press. doi:10.1016/b978-0-12-813499-2.00004-5