Topic Modelling of India's Digital Healthcare Research Trend

Munikrishnappa Anilkumar, Manasa Nagabhushanam, Mallieswari R

M S Ramaiah Institute of Management, Bengaluru, India

Corresponding author: Munikrishnappa Anilkumar, Email: anilkumar@msrim.org

New advancements in information technology have had a tremendous impact on digital healthcare applications in the medical area. The study themes connected to digital healthcare technology and its intervention must be discovered and studied systematically. As a research gap, digital healthcare research in India has yet to be investigated thematically using topic modeling; in this context, the study employs topic modeling's Non-Negative Matrix Factorization algorithm to systematically generate digital health research themes in India. After preprocessing, the raw texts were transformed into Term Frequency-Inverse Document Frequency vectors. The Non-Negative Matrix Factorization approach from topic modeling was used for text classification. The k parameter was used for feature selection, yielding a set number of topics for semantic interpretation. Analysis of the research articles revealed that there has been considerable growth in digital healthcare research in India since 2017; the majority of publications occurred in 2020 and 2021, with less previous to 2017. Topic modeling of 97 published articles yielded the top three research themes: evaluation, public policy, and communities. The findings from the research themes will provide a thematic understanding of digital healthcare research in India. It will also aid future studies through text analysis, topic modeling, and decision-making in digital healthcare treatments.

Keywords: Digital Healthcare, Topic Modelling, Public Policy, Text Analysis.

2024. In Mukesh Saraswat & Rajani Kumari (eds.), *Applied Intelligence and Computing*, 169–176. Computing & Intelligent Systems, SCRS, India. DOI: https://doi.org/10.56155/978-81-955020-9-7-18

1 Introduction

Healthcare information systems are an impetus in promoting and expanding digital healthcare in India. The National Health Policy of 1983 and 2002 has played a pivotal role in providing the base for various strategies to achieve its goals and objectives. As policies around public healthcare moved in a new direction of comprehensiveness, its approach focused on adopting innovative digital healthcare systems across the country. The pathways related to research in digital health are visible through various policies and program initiatives starting with National Health policy of 2017 (NHP 2017).

Various studies have been done to understanding the digital health care research using topic modelling. The digital health intervention literature review related to hypertension patients[1]; A literature review to understand digital health intervention during COVID response was done[2]; A digital healthcare innovation ecosystems literature review and suggested a conceptual framework[3]; Another disease specific literature study has been the digital health intervention for mental illness[4]. The studies so far involving the digital health literature are not comprehensive and India specific study, they are mainly a disease specific, and there are no studies using Non-Negative Matrix Factorization algorithm to generate topics with respect to India.

This study based on the topic modeling shows that digital healthcare research themes of India can be categorized into three main themes i.e., Evaluation; Public Policy; Communities. Below method and result sections will explain the process for generating themes, and further engages in thematic discussion on digital healthcare in India.

2 Conceptual Framework

The published research articles in English language required for the analysis were downloaded from the Scopus using advanced search strategy. Based on the discussion amongst the authors and iterating multiple search strings, the string chosen was TITLE-ABS-KEY ("digital health" AND "India") which generated 97 articles as on 28th March 2022 for the period 2014-2022.

The unstructured text corpora in the form of abstracts were systematically analyzed using the Natural Language Processing's (NLP) topic modelling method. The topic modelling algorithm used for generating the features or topics is Non-Negative Matrix Factorization (NMF) algorithm developed in 1999 [5]. NMF matrix factorization technique is widely used amongst researchers for topic generation [6]. An NMF "for a non-negative data matrix V, it finds an approximate factorization $V \approx WH$ into non-negative factors W and H" [7]. Though there are other topic modelling algorithms, NMF is better suited for smaller dataset with words having better coherence. The NMF python library package [8] and python coding was done on Jupyter notebook to generate the topics on the digital health research publications of India. To generate a better semantically interpretable number of topics, a k parameter with CUMass intrinsic coherence measure [9] was adopted in the study. The number of topics were generated based on the k parameter, a best topic having better coherence were defined and interpreted.

Before the application of NMF algorithm, a step wise pre-processing techniques were conducted on the text which included tokenization, lemmatization, removal of stop words using machine learning library scikit-learn [10] and Natural Language Toolkit [11], and further pronouns, and punctuations were removed. The title and key words contain frequently occurring words, so they were excluded from analysis to get a better coherence. Before using NMF algorithm the pre-processed text, here the abstracts, were converted into a 'log-based Term Frequency-Inverse Document Frequency (TF-IDF) vectors' [12].

Topic modelling of unstructured text has its own limitations as the topics are not explicit or meaningful, they are just the bag of words. In the study, interpreting the topics having bag of words was left to the authors subjectivity. As much as possible the subjectivity in interpretation was contained by

verifying the topics between the authors. The full extent of understanding the themes on digital health research has limitation as the publications downloaded were limited to Scopus indexed journals and the data post 28th March 2022 is considered due to non-accessibility issue.

3 Analysis and Results

Out of parameter k = 30 topics generated from the research abstracts, the k=3 topics were found to be having better coherence (Figure 1) and semantically interpretable. The best three automatically generated topics were 0,1, and 2 which are mentioned below.

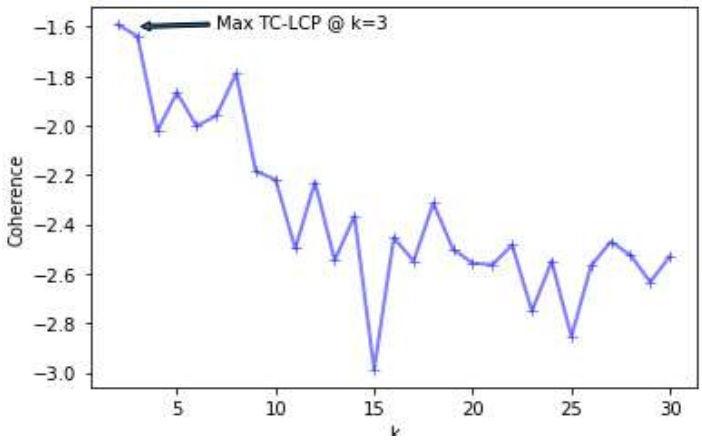

Figure 1. Coherence value and topics

Topic #0: ['care', 'data', 'system', 'mobile', 'service', 'information, 'provide', 'develop', 'technology', 'program', 'primary', 'design', 'base', 'train', 'model', 'quality', 'phone', 'national', 'increase', 'clinical, 'delivery', 'study', 'medical', 'patient', 'scale', 'support', 'treatment', 'across', 'systems', 'learn', 'paper', 'record', 'platform', 'time', 'enable', 'primary care', 'low', 'surveillance', 'include', 'worke', 'level', 'practice', 'state', 'cost', 'analysis', 'implementation', 'challenge', 'context', 'make', 'monitor', 'literacy', 'infrastructure', 'propose', 'fit', 'patients', 'need', 'initiatives', 'solutions', 'number', 'care provide']

Topic #1: ['covid', 'care', 'pandemic', 'new', 'covid pandemic', 'coronavirus', 'country', 'technologies', 'world', 'market', 'population', 'public', 'doctor', 'artificial intelligence', 'intelligence', 'artificial', 'telemedicine', 'still', 'growth', 'offer', 'big', 'government', 'year', 'innovation', 'role', 'thus', 'apps', 'utilize', 'combat', 'platforms', 'connect', 'along', 'various', 'post', 'media', 'development', 'relate', 'issue', 'potential', 'article', 'response', 'medical', 'need', 'high', 'disease', 'care system', 'opportunity', 'future', 'make', 'big data', 'highlight', 'outbreak', 'service', 'remote', 'case', 'demand', 'spread', 'also', 'major', 'insights']

Topic #2: ['study', 'research', 'women', 'interventions', 'risk', 'community', 'intervention', 'tool', 'review', 'access', 'result', 'find', 'base', 'article', 'outcomes', 'include', 'technologies', 'search', 'participants', 'influence', 'impact``', 'publish', 'conduct', 'south', 'identify', 'experience', 'six', 'assess', 'improve', 'evidence', 'disease', 'self', 'aim', 'adherence', 'group', 'factor', 'online', 'unite', 'focus', 'undetand', 'mental', 'social', 'objective', 'track', 'management', 'three', 'global', 'estimate', 'assessment', 'literature', 'methods', 'significant', 'status', 'age', 'people', 'policy', 'live', 'report', 'background', 'purpose']

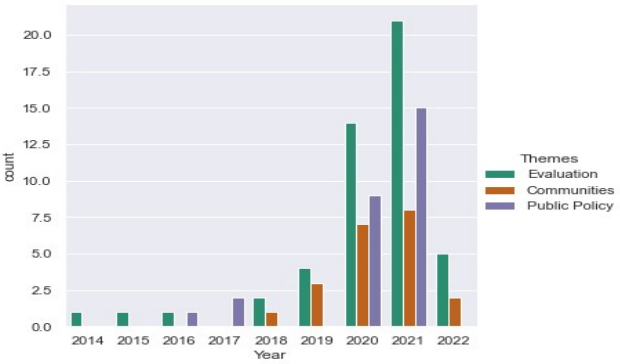

Figure 2. Year-wise themes

Based on the interpretability, the three topics above were coded to rename it as 'Evaluation', 'Public Policy', and 'Communities', they are the three broad research themes on digital health care of India over a nine-year period. Analysis of number of publications (Figure 2) shows that, there is an increasing publication of digital healthcare from India from the year 2018 onwards. Most of these publications have happened between the years 2020 and 2021, and limited number of articles published prior to 2017. As the search for 2022 to 28th March of the year, there is a low publication, but there is an indication it might see increase in those numbers.

As Figure 3 shows, the dominant research themes being the 'Evaluation' of digital healthcare systems and technologies, accounting for 50.5% of the total publications. It is the most significantly researched area over nine-year period starting from 2014. Second most popular theme being 'Public Policies' of the government related to digital healthcare accounting 27.8% of the total publications. Finally, 'Communities' in relation to digital healthcare accounting 21.6% of the total publications.

Though research articles on evaluating the digital health technologies increases from the year 2019, it peaks in the year 2020 and 2021. Research on public policies of digital healthcare sees increase on 2020 and 2021, the two years effected by COVID pandemic. Similarly, communities which represent the providers and beneficiaries of digital health sees the increase in publications during 2020 and 2021.

4 Discussion

As the policy ecosystem is promotive of increasing the research publications, the digital health area is seeing momentum in that direction. The analysis shows that since National Health Policy 2017 there is an overall increase in the number of research knowledge on digital health generated through papers published in peer reviewed journals (Figure 3).

4.1 Evaluation of Digital Healthcare Systems and Technologies

The 'Evaluation' research theme broadly focuses on evaluating the different digital technologies, applications, and healthcare informatics systems, and architecture on a sample population. The diverse stakeholder's perspective on digital healthcare is critical for the successful and effective utilization of technologies. In this regards, there is positive response to the efficacy of Integrated

Health Information System for Primary Health Care (IHIS4PHC) in Chandigarh, though positively perceived, there is a concern regarding transparency, accountability and digital literacy [13]. Partnerships between the government and private is crucial for an effective digital health intervention, which could be achieved by public private partnerships (PPP). The 'Technology enabled Remote Health Care' using public private partnerships is totally possible with the help of dedicated e-health teams in executing the digital health projects [14]. Healthcare infrastructures provides a much needed support for digital health technologies for their effective interventions, these infrastructures are lagging in the 'lower and mid-tier healthcare facilities compared with apex facilities in India' [15].

The new innovations and developments in the domain of Artificial Intelligence (AI), and Machine Learning (ML) techniques are revolutionizing the digital health technologies in diagnosing and treating the patients. AI & ML efficiency is being studied in India through various projects, SnehAI is one such AI enabled chat bot developed by Population Foundation of India, which has effectively used NLP techniques to educate adults, and promote sexual and reproductive health [16]. Further, data generated through health technology applications needs to undergo regular quality checks, which needs to be regularly monitored using data analytics and machine learning techniques [17].

4.2 Public Policy and Digital Health Care

Primarily, 'Public Policy' research is related to the policies dealing with using technology to contain COVID-19 pandemic, there are few policy studies on technology related to non-COVID aspects. The innovative tools of digital healthcare systems played a vital role globally and in India to address the medical crisis due to COVID-19 pandemic. These technologies were used for the purpose of surveillance, tracking patients, use of telehealth interventions, as diagnostic support, support for healthcare workers and also for administrative support [18]. The policies and strategies around digital health interventions during these pandemics was well studied. As the year 2020 and 2021 were very critical for the government to deal with pandemic, it significantly used the digital health means to face the healthcare crisis.

The themes on public policy research largely focuses on governments policies and programs related digital healthcare. The leveraging of technology to combat COVID-19 using the policies such as National Digital Health Mission and Atma Nirbhar Bharat Scheme were emphasized in the research [19]. As the digital health systems were used by the government to address the a vaccine hesitancy amongst the public, information dissemination through 'official digital platforms' were considered critical and were studied during rise in COVID cases [20]. A quality medical care is essential for a quality health, providing such care is much easier in urban set up compared to rural areas due to gaps in healthcare infrastructure, this is where an effective use of Telemedicine becomes critical especially during pandemic. India's 'national patient-to-doctor' telemedicine service 'eSanjeevaniOPD' were expanded to address pandemic challenges [21].

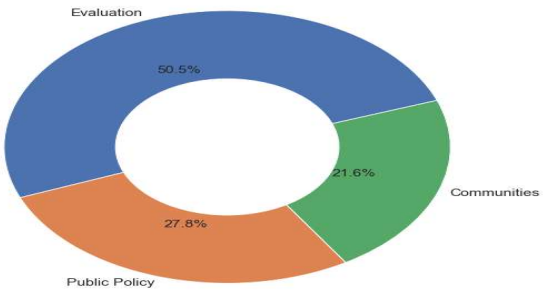

Figure 3. Digital health research themes of India

There are some studies on policies, technology, and health related to non-covid comes under the public policy theme. The research themes emphasized were related to digital health policies concerning to data privacy and digital divide. These policy challenges could affect the diabetes care in the country, hence, research calls for a 'strong technology adaptation and policy interventions' for 'digitalization of diabetes care' [22]. The health startups are critical for digital expansion in the country, the policies and programs are essential for promoting the health startups [23]. The data generated from the digital health systems are tremendous, which falls into the category of Big data, the impact of Big Data on the healthcare systems are were critically studied during [24].

4.3 Communities and Digital HealthCare in the Future

The other published research was related to future studies on digital health and communities. 'Communities', broadly represent healthcare providers and receivers, includes care givers, old age people, adults, pregnant and post-natal women, and gender perspectives in general. There are specific health issues which demands a specialized training and support to caregivers especially for a dementia effected patients which will have prevalence of 150 million by 2050, the caregivers are assisted by 'iSupport' platform of World Health Organization [25]. The increasing prevalence of mobile devices in healthcare is evident, as they assist healthcare providers in making clinical decisions. However, the studies show that there is an uncertainty over the effectiveness of using mobile devices for clinical decisions on specific health areas [26].

The healthcare sector is seeing technological advancements in providing much-needed support to vulnerable groups such as the elderly. In this regard, the 'mental health and digital health services' are the primary influencers on the aged population during COVID-19 [27]. Digital health care is also increasingly focusing on diagnosing and treating pregnant and post-natal women. Technological interventions for pre- and post-natal women in India need to use 'appropriate technology,' which is multi-lingual, to encourage information-seeking behavior [28].

5 Conclusion

Text classification using e-health information has numerous applications in the medicinal field. For such classifications, it is crucial to interpret categorization. Utilizing topic modeling, such as topic embedding's, for text classification can enhance a model's interpretability. The study systematically applied topic modeling techniques using the NMF algorithm to study the literature on digital healthcare. The results helped generate the three major themes, 'Evaluations,' 'Public Policy, ' and 'Communities,' and digital health interventions are happening at these three levels. The analysis also showed the increasing trend for digital healthcare interventions since the pandemic. The NMF model, if further trained with a large dataset, can be used for predictive purposes, which would help medical practitioners, the health sector, academicians, and public policymakers to update themselves on the current digital health research themes.

References

[1] K. Wechkunanukul, D. R. Parajuli, and M. Hamiduzzaman, "Utilising digital health to improve medication-related quality of care for hypertensive patients: An integrative literature review," World J. Clin. Cases, vol. 8, no. 11, pp. 2266–2279, 2020, doi: 10.12998/wjcc.v8.i11.2266.

[2] "Assessing the Implementation of Digital Innovations in Response to the COVID-19 Pandemic to Address Key Public Health Functions: Scoping Review of Academic and Nonacademic Literature," JMIR Public Health Surveill., vol. 8, no. 7, 2022, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85134361100&doi=10.2196%2f34605&partnerID=40&md5=606c8b7a36278fb68e7d23b80de04d46

[3] G. E. Iyawa, M. Herselman, and A. Botha, "Digital Health Innovation Ecosystems: From Systematic Literature Review to Conceptual Framework," Procedia Comput. Sci., vol. 100, pp. 244–252, Jan. 2016, doi: 10.1016/j.procs.2016.09.149.

[4] S. Batra, R. A. Baker, T. Wang, F. Forma, F. DiBiasi, and T. Peters-Strickland, "Digital health technology for use in patients with serious mental illness: a systematic review of the literature," Med. Devices Evid. Res., vol. 10, pp. 237–251, Dec. 2017, doi: 10.2147/MDER.S144158.

[5] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, no. 6755, Art. no. 6755, Oct. 1999, doi: 10.1038/44565.

[6] S. Arora et al., "A Practical Algorithm for Topic Modeling with Provable Guarantees," arXiv, arXiv:1212.4777, Dec. 2012. doi: 10.48550/arXiv.1212.4777.

[7] P. O. Hoyer and P. Hoyer, "Non-negative Matrix Factorization with Sparseness Constraints," J. Mach. Learn. Res., p. 13, 2004.

[8] "sklearn.decomposition.NMF," scikit-learn. Accessed: May 24, 2022. [Online]. Available: https://scikit-learn/stable/modules/generated/sklearn.decomposition.NMF.html

[9] D. Mimno, H. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing Semantic Coherence in Topic Models," in Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK: Association for Computational Linguistics, 2011, pp. 262–272.

[10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., p. 6, 2011.

[11] E. Loper and S. Bird, "NLTK: the Natural Language Toolkit," in Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics - Volume 1, in ETMTNLP '02. USA: Association for Computational Linguistics, Jul. 2002, pp. 63–70. doi: 10.3115/1118108.1118117.

[12] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Inf. Process. Manag., vol. 24, no. 5, pp. 513–523, Jan. 1988, doi: 10.1016/0306-4573(88)90021-0.

[13] D. S. Faujdar, T. Singh, M. Kaur, S. Sahay, and R. Kumar, "Stakeholders' Perceptions of the Implementation of a Patient-Centric Digital Health Application for Primary Healthcare in India," Healthc. Inform. Res., vol. 27, no. 4, pp. 315–324, Oct. 2021, doi: 10.4258/hir.2021.27.4.315.

[14] K. Ganapathy et al., "Digital Health Care in Public Private Partnership Mode," Telemed. J. E-Health Off. J. Am. Telemed. Assoc., vol. 27, no. 12, pp. 1363–1371, Dec. 2021, doi: 10.1089/tmj.2020.0499.

[15] S. A. Patel et al., "A model for national assessment of barriers for implementing digital technology interventions to improve hypertension management in the public health care system in India," BMC Health Serv. Res., vol. 21, no. 1, p. 1101, Oct. 2021, doi: 10.1186/s12913-021-06999-9.

[16] H. Wang et al., "An Artificial Intelligence Chatbot for Young People's Sexual and Reproductive Health in India (SnehAI): Instrumental Case Study," J. Med. Internet Res., vol. 24, no. 1, p. e29969, Jan. 2022, doi: 10.2196/29969.

[17] N. Shah et al., "SMS feedback system as a quality assurance mechanism: experience from a household survey in rural India," BMJ Glob. Health, vol. 6, no. Suppl 5, p. e005287, Jul. 2021, doi: 10.1136/bmjgh-2021-005287.

[18] A. Kapoor, S. Guha, M. Kanti Das, K. C. Goswami, and R. Yadav, "Digital healthcare: The only solution for better healthcare during COVID-19 pandemic?," Indian Heart J., vol. 72, no. 2, pp. 61–64, Mar. 2020, doi: 10.1016/j.ihj.2020.04.001.

[19] A. Ramachandran and S. N. Sarbadhikari, "Digital Health for the post-COVID-19 Pandemic in India: Emerging Technologies for Healthcare," in 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), Mar. 2021, pp. 244–249.

[20] R. Kaur, A. Jain, and J. Sharma, "EFFECTIVENESS OF HEALTH RISK COMMUNICATION DURING PANDEMIC: AN EXPLORATIVE STUDY," J. Content Community Commun., pp. 176–187, 2021, Accessed: May 17, 2022. [Online]. Available: https://doi.org/10.31620/JCCC.12.21/14

[21] C. Naithani, S. P. Sood, and A. Agrahari, "The Indian healthcare system turns to digital health: eSanjeevaniOPD as a national telemedicine service," J. Inf. Technol. Teach. Cases, p. 20438869211061575, Dec. 2021, doi: 10.1177/20438869211061575.

[22] J. Kesavadev, G. Krishnan, and V. Mohan, "Digital health and diabetes: experience from India," Ther. Adv. Endocrinol. Metab., vol. 12, p. 20420188211054676, 2021, doi: 10.1177/20420188211054676.

[23] L. Motha, R. Nalini, A. R., R. Amudha, and V. Badrinath, "Health Startups in India-A Progression Towards Development," Res. J. Pharm. Technol., vol. 10, pp. 4175–4177, Dec. 2017, doi: 10.5958/0974-360X.2017.00761.2.

[24] R. Duggal, S. Balvinder, and S. K. Khatri, "Opportunities and Challenges of Using Big Data Analytics in Indian Healthcare System," Indian J. Public Health Res. Dev., vol. 7, p. 238, Oct. 2016, doi: 10.5958/0976-5506.2016.00226.6.

[25] T. A. Nguyen et al., "Empowering Dementia Carers With an iSupport Virtual Assistant (e-DiVA) in Asia-Pacific Regional Countries: Protocol for a Pilot Multisite Randomized Controlled Trial," JMIR Res. Protoc., vol. 10, no. 11, p. e33572, Nov. 2021, doi: 10.2196/33572.

[26] S. Agarwal et al., "Decision-support tools via mobile devices to improve quality of care in primary healthcare settings," Cochrane Database Syst. Rev., vol. 7, p. CD012944, Jul. 2021, doi: 10.1002/14651858.CD012944.pub2.

[27] P. Bastani, M. Mohammadpour, M. Samadbeik, M. Bastani, G. Rossi-Fedele, and M. Balasubramanian, "Factors influencing access and utilization of health services among older people during the COVID -□19 pandemic: a scoping review," Arch. Public Health Arch. Belg. Sante Publique, vol. 79, no. 1, p. 190, Nov. 2021, doi: 10.1186/s13690-021-00719-9.

[28] A. Joshi, D. Roy, A. Ganju, M. Joshi, and S. Sharma, "ICT Acceptance for Information Seeking Amongst Pre- and Postnatal Women in Urban Slums," in Human-Computer Interaction – INTERACT 2019, D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, and P. Zaphiris, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 152–160. doi: 10.1007/978-3-030-29387-1_9.