

Health Misinformation Vs. Facts on Social Media: Co-Occurrence Network Analysis in Bangladesh

Parinda Rahman and Ifeoma Adaji

University of British Columbia, Kelowna, Canada

parinda1@student.ubc.ca

ifeoma.adaji@ubc.ca

Abstract: The increased usage of social media provides a way to disseminate health-related information more quickly. Alternatively, sharing health content on social media poses risks due to unrestricted posting, enabling misinformation to spread. Regional social and cultural contexts influence themes in social media posts, underscoring the importance of understanding content and prevalent misinformation themes. This insight is crucial for tailoring interventions, resource allocation, misinformation detection algorithms, and policy formulation. We conducted word co-occurrence network analysis, creating and analyzing two networks for valid information and misinformation in Bangladesh. The prevalence of misinformation regarding natural ingredients and treatments in Bangladesh underscores the need for targeted efforts to combat health misinformation on social media. For each network, we computed metrics such as betweenness, Katz centrality, out-degree, and degree distribution. Furthermore, we computed the Louvain clustering algorithm to identify word clusters. A comparative analysis of both networks suggested that the context of words used in sentences was important and that both networks contained information about natural remedies or ingredients for health benefits. The misinformation network contained the word *raw turmeric* with the highest bigram frequency of 162. These natural remedies were stated as cures, and there was much misinformation and valid information surrounding common health conditions such as blood pressure. This was depicted through the word *blood* having an outdegree of four and seven in the misinformation and valid information networks, respectively. The valid information network emphasized the beneficial properties of natural ingredients rather than their supposed ability to cure diseases. This study provides insights into the distinctions and parallels between valid health information and misinformation on social media, considering their social and cultural context. It underscores shared semantics and bigram words between them, suggesting that understanding these differences can aid in addressing region-specific challenges.

Keywords: Health, Social media, Misinformation, Social network analysis, Word co-occurrence Network

1. Introduction

Social media serves as a prominent channel for health information dissemination (Chou, Oh and Klein, 2018), utilized by organizations like the World Health Organization (Xiong and Liu, 2014). However, it also poses unprecedented risks as it allows any user to spread health-related information (Cavallo et al., 2014). Misinformation tends to proliferate over social media more than valid scientific information (Vosoughi, Roy and Aral, 2018). In literature, an assertion that is misleading and lacks scientific backing surrounding human health can be referred to as health misinformation (Funk, Salathé and Jansen, 2010). Health misinformation, lacking scientific support, surged during the COVID-19 pandemic, with false cures and rumors circulating widely (Grimes, 2021). Examples include claims of curing COVID-19 with chloroquine, cow urine, alcohol, and hot water (Mackey et al., 2021). Apart from the pandemic, numerous self-proclaimed health experts and practitioners of alternative medicine have promoted unproven medications, concoctions, suggestions, and treatments as means of "boosting" the immune system over social media (Caulfield, 2020). This can lead to harmful outcomes such as deaths from alcohol poisoning from concoctions (Trew, 2020). However, much of the misinformation is motivated by cultural practices, societal stigma about certain diseases, and locally available ingredients (Wang, 2018). While the problem of health misinformation is global, in recent years between 2021 and 2022, Bangladesh has observed a 10.1% rise in its social media user base (Simon, 2021). While much research on this topic focuses on Western countries due to data availability, it's noted that the nature and spread of health misinformation depend heavily on social and cultural contexts (Wang, 2018). Bangladesh's dense population suggests that misinformation could affect many people (Khan et al., 2021).

Past studies have utilized machine learning and deep learning techniques for detecting misinformation (Cabral et al., 2021) but have largely overlooked analyzing themes present in the different contexts between misinformation and valid information. However, customizing algorithms according to regional misinformation contexts could enhance detection and inform targeted policy interventions. In the health misinformation domain, there is little to no literature that analyses the context of misinformation in a specific region or Bangladesh. A recent study focused on COVID-19-related misinformation in India analyzed the themes present and their context using co-occurrence network analysis using bigrams (Naeem, Bhatti and Khan, 2021). However, that work was limited to COVID-19-related misinformation. Co-occurrence network analysis using

bigrams encapsulates context by examining the frequency of items appearing together within a defined context and discerning relationships between consecutive word pairs making it suitable for studying health misinformation on social media (Hu, Huang, & Wang, 2018; Nistor & Zadobrischi, 2022).

Addressing this gap, this research aims to conduct a comparative analysis of content between health misinformation and valid information in Bangladesh by exploring bigram co-occurrence network analysis technique. The work aims to investigate what themes of misinformation are prevalent and whether there are differences or similarities with valid information. The work focuses on a comprehensive analysis across diverse health topics as the majority of current literature focuses on a specific disease or on COVID-19 which limits the scope of application. The novel contributions of our work are:

- Curating text data sourced from Facebook through manual scraping.
- Conducting bigram network visualization, community detection, and metric analysis (in-degree, out-degree, centrality measures) on Bangladesh-specific health information.
- Comparing networks of health misinformation and valid information to gain insights.

2. Literature Review

2.1 Health Misinformation on Social Media

The influence of misinformation on social media poses a significant challenge to public health, potentially undermining the efficacy of programs, campaigns, and initiatives designed to promote citizen health, awareness, and well-being (Pulido et al., 2020). Wang et al (2018) conducted a systematic literature review and found that the type of health information shared is often driven by social and cultural contexts. Another study shows the prevalence of misinformation surrounding family planning and contraception in Nigeria due to the lack of awareness about reproductive health and religious teachings (Ankomah, 2011). Sentell et al (2020) identified misinformation surrounding vaping was present in Hawaii due to the misappropriation of native culture. The study underscored the need for customized health interventions targeting specific communities and cultures to counteract the compelling nature of health misinformation online, particularly concerning its impact on youth and the perpetuation of health disparities. These works provide evidence that misinformation is largely dependent on regional social and cultural context.

2.2 Misinformation in South Asia

Previous literature identifies that in Bangladesh, India, and Pakistan, religious beliefs significantly increased vaccine hesitancy (Kanozia and Arya, 2021). Moreover, health practices in this region are guided by many ancient practices of Ayurveda (Haque et al., 2018); thus, a lot of health practices focus on the use of natural ingredients. Ahmed et al (2020) identified that during the COVID-19 pandemic in Bangladesh, preventive measures and beneficial properties of natural ingredients were misrepresented as cures. Moreover, previous studies also suggest that severe societal stigma, a lack of sexual health education, and misinformation surrounding sexual health are prevalent in South Asia (Banik et al., 2023). This work can be used to investigate whether these themes of misinformation are prominent in social media specifically in Bangladesh.

2.3 Network analysis and Co-Occurrence Networks

Previous studies have shown the efficacy of network analysis in understanding the context of health-related misinformation. For instance, one study focused on HPV vaccine misinformation on Instagram, where separate networks for pro and anti-vaccine posts were created, revealing features such as conspiracy theories and unsupported assertions (Massey et al., 2020). Another study during the Zika pandemic examined structural differences between networks disseminating misinformation and accurate information, using metrics like modularity for comparison (Safarnejad et al., 2020). The study compared nine metrics, such as modularity, out-degree, etc., to understand the difference between the two network behaviors. Bigram analysis within network analysis has been widely used to analyze text data. It allows for the examination of lexical structure and semantics, as seen in studies investigating sentiment in political tweets and co-occurring words in tweets discussing vaccine-related blood clots (Fudolig et al., 2022). The majority of these studies focused on a particular health domain such as vaccination or Zika virus, limiting the investigation of other misinformation contexts.

Hence, prior research underscores the significance of cultural and societal contexts in the realm of health-related misinformation. However, despite the prevalence of such misinformation in South Asia, there is a notable absence of studies specifically targeting Bangladesh and comprehensively exploring the contextual

nuances across various health domains using network analysis metrics. Existing studies employing co-occurrence analysis tend to concentrate on singular health domains, highlighting the importance and novelty of this work.

3. Methodology

3.1 Data Collection

Due to Facebook not having a publicly available application programming interface (API), automated data collection was not feasible. Bangla text data was manually gathered from publicly accessible Facebook posts by searching using relevant Bangla keywords as no publicly available dataset on health information in Bangladesh existed. The Facebook posts originated between the time frame of 2018 to 2022. The list of keywords used for the data collection was “health”, “cure”, and “disease”. Specific disease-related terms were not considered as keywords for the search rather general words were considered. Facebook, being the most widely used platform for information access in Bangladesh, was chosen as the data source (Al-Zaman, Md Sayeed and Noman, 2021). The collected data was manually classified as false or valid information. The data was translated using Google Translate. To ensure semantic accuracy three native Bengali speakers validated and modified the translations wherever it was required. Data labeling was further validated by a group of Bangladeshi medical specialists proficient in Bengali. English translations of the data were used for the analysis. A total of 3266 credible information posts and 2719 misinformation posts were collected. Figure 1 shows an example data point from the dataset.

Text	Translation
ধূমপান, মদ্যপান এবং অন্যান্য ক্ষতিকারক জীবনযাত্রার কারণগুলি ক্যান্সারের ঝুঁকি বাড়াতে পারে না, বরং এটি কেবল দুর্ভাগ্য হতে পারে গবেষণায় দেখা গেছে	Smoking, Drinking and other harmful lifestyle factors does not increase the risk of cancer, rather it simply may be bad luck study finds

Figure 1: Example from the data set of the original Bangla text of the Facebook post and its translation

3.2 Data Preprocessing

Text data segments were concatenated and processed in R using the "tm" library. It was converted to lowercase, and numbers, punctuation, and special characters were also removed. Stop words were removed to retain only significant words. The cleaned text was converted to a data frame, and bigrams were generated separately for valid information and health misinformation using "tidyr". Bigrams with an occurrence frequency of 20 or more were selected for analysis. This criterion was chosen because recurrent bigrams signify the prevalence of word pairs (Evert, 2005). Even after this filtering, over 85% of the total number of bigrams was retained in the network, indicating significant representativeness of the dataset. Following filtration, visualizations were generated using the igraph package in R.

3.3 Data and Network Analysis

For analysis and visualization, two separate directed networks for misinformation and valid information were created, with words representing the nodes and edges indicating connections between two words if they appear one after another in sequence. If a word is pointed toward, it signifies that the word is the second word in the bigram. For example, if in the visualization the word “blood” has a directed edge towards the word “pressure”, it is the second word. Edge thickness represents bigram frequency. The following metrics were computed on both networks to gain insights about their content.

3.3.1 Bigram frequency

This metric represents the number of times each bigram occurs in the entire document. Bigrams that occur frequently signify significant and commonly used word pairs in each network, offering insights into the prevalent themes within the dataset.

3.3.2 In-degree

The adjacency matrix's i^{th} row sum equals the in-degree of node i , which represents the total number of connections onto node i (Newman, 2018). In the context of this analysis, the in-degree specifically indicates words that appear as the second word in the bigram pairs within the network.

3.3.3 Out-degree

The entire number of connections originating from node i is represented by the out-degree of node i , which is the sum of the i^{th} column in the adjacency matrix (Newman, 2018). The out-degree is crucial as it indicates how many other words a particular word connects to within the network.

3.3.4 Betweenness

The betweenness centrality provides information about the flow of information in the network (Newman, 2018). The flow of connector words in the network could be observed to gain insights into the themes and use of language.

3.3.5 Katz Centrality

Katz centrality calculates the relative importance of a node within a directed network by counting the number of immediate neighbors (first-degree nodes) and all other nodes in the network that connect to the node under consideration through these immediate neighbours (Newman, 2018). This measure is crucial in directed networks as it provides insight into the relative importance and influence of words, considering both direct connections and indirect ones.

3.3.6 Louvain Clustering

This community detection method partitions networks into densely connected clusters to optimize modularity, a measure of network division strength. Selected for its speed and availability in open-source libraries, the Louvain algorithm can also detect hierarchical structures within words, revealing prevalent themes (Newman, 2018) and it is widely used for directed networks.

4. Findings

4.1 Basic Network Characteristics

4.1.1 Bigram frequency and network visualization

Table 1 shows the top 5 bigram frequencies in the valid information network and the misinformation network.

Table 1: Top 5 bigram frequencies in the misinformation and the valid information network

Misinformation network			Valid information network		
Word 1	Word 2	Frequency	Word 1	Word 2	Frequency
raw	turmeric	162	blood	pressure	228
black	cumin	128	neem	leaves	125
blood	pressure	108	lemon	juice	121
hot	water	72	blood	sugar	104
leaf	juice	68	heart	disease	76

The misinformation network words from Table 1 highlight that "raw turmeric" and "black cumin" were the most frequently occurring words in the misinformation network, often suggested as remedies for various diseases. Similarly, "leaf juice" and "hot water" indicate the prevalence of natural ingredients as treatments. This aligns with previous findings in India, where garlic and hot water were influential nodes in similar network analyses (Naeem, Bhatti and Khan, 2021). Additionally, the presence of "blood pressure" suggests misinformation surrounding common diseases, consistent with the challenge of misinformation in noncommunicable diseases like hypertension among Bangladeshis (Jafar et al., 2018). The term "blood pressure" appears most frequently in the valid information network. Upon examining the dataset, it was discovered that words like "neem leaves" and "lemon juice" were used to describe the health benefits of these ingredients rather than promoting them as cures, as observed in the misinformation network. Additionally, in the valid information network, words such as "blood," "hair," and "juice" remained popular nodes. Given the presence of medicinal healers utilizing natural ingredients in the country (Rahmatullah et al., 2010; Jahan, 2010), discussions about the beneficial properties of these ingredients are prevalent in the valid information network. Figure 1 shows the misinformation network and Figure 2 shows the communities in the network after using the Louvain algorithm. Due to network disconnection, most communities are small in size, with 24 identified in total, the largest containing 9 nodes. Prominent communities formed around words like "hair,"

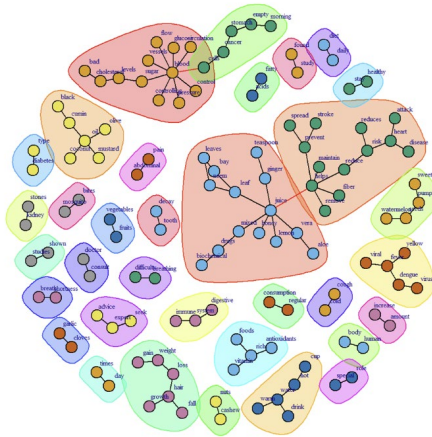


Figure 4: Community detection using the Louvain algorithm for the valid information network

4.2 In-Degree

Table 3 presents the nodes with the highest in-degree for both the misinformation and valid information networks. Nodes with a higher in-degree represent the second word in the bigram. Notably, both networks feature overlapping words such as "system," "juice," and "oil," suggesting similar lexical and semantic usage. However, the contextual meaning of these words varies significantly. This aligns with Garg and Kumar, (2018), which suggests that word co-occurrence networks can be context-dependent, as observed in this paper.

Table 3: Top 5 In-degree of each network

Misinformation Network		Valid Information Network	
Word	Indegree	Word	Indegree
system	3	juice	4
juice	2	oil	4
oil	2	fever	3
levels	2	blood	2
loss	2	helps	2

4.3 Out-Degree

Table 3 lists the top 5 out-degrees of nodes in each network, representing influential nodes that point to the second word in a bigram. Notably, the misinformation network uniquely features words like "herbal" and "thankuni" (Centella Asiatica Plant), indicating the prevalence of natural cures and alternative medicine, consistent with Hossain and Haque (2023), highlighting the false spread of "thankuni" leaves as a COVID-19 cure in Bangladesh. Additionally, "blood" emerges as an influential node in both networks, suggesting the prevalence of discussions surrounding blood-related diseases in social media health discussions. Furthermore, the presence of "hair" in both networks indicates a popular topic in health information, consistent with Iglesias-Puzas et al (2021) identifying dermatology, including hair-related discussions, as a prominent area on social media, often containing misleading cures.

Table 3: Top 5 Out-degree of each network

Misinformation Network		Valid Information Network	
Word	Out degree	Word	Out degree
blood	4	blood	7
hair	4	helps	4
hot	3	hair	3
thankuni	2	heart	2
herbal	2	dengue	2

4.4 Centrality Measures

4.4.1 Betweenness

Words with high bigram frequency as seen in Table 1 were also reported to have a high betweenness score. However, due to numerous structural holes or disconnected components in both networks, it may not fully capture node influence (Hassanzadeh, Khodadust and Zandian, 2012). In the misinformation network, "turmeric" and "raw" had the highest betweenness scores (15 and 12, respectively). Conversely, in the valid information network, "juice" and "helps" scored 104 and 100, respectively. Therefore, this indicates that in Bangladesh "raw turmeric" reported as a cure in the misinformation network. The presence of the word "helps" as an influential node in the valid information network suggests that the benefits of certain substances were highlighted as opposed to stating it as a cure.

4.4.2 Katz centrality

Table 2 displays the Katz centrality scores for each network. Interestingly, both networks exhibit an overlap of nodes with high Katz centrality scores, suggesting that similar words hold influence in both networks. Consequently, distinguishing between the two networks without sentence context proves challenging. For example, the word "oil" and "juice" reported high Katz centrality values. Therefore, no insight could be drawn from the nodes and their Katz centrality score without additional context.

Table 2: Top 5 Katz centrality of each network

Misinformation Network		Valid Information Network	
Word	Katz Centrality	Word	Katz Centrality
system	1.30	juice	1.44
oil	1.22	oil	1.41
levels	1.21	fever	1.30
juice	1.21	mixed	1.25
loss	1.20	helps	1.24

4.5 Comparative Analysis

Results indicate that lexical and semantic features alone are insufficient to differentiate between misinformation and valid information. Contextual understanding is vital, as many nodes share significant metrics across both networks. For example, the word "blood" displays contextual differences: in the valid information network, it has an out-degree of 7, while in the misinformation network, it has an out-degree of 4. While both networks touch upon themes of natural remedies and Ayurveda for health, a notable distinction lies in their portrayal. In the misinformation network, natural remedies are often touted as cures, whereas the valid information network emphasizes their role as preventive measures, as evidenced by terms like "helps" and "reduces." See Figure 5 for the contextual variations of the word "blood."

MISINFORMATION	VALID INFORMATION
People with low blood sugar levels should never eat too much cloves. Because, in their case, consumption of cloves can cause hyperglycemia.	Obesity increases the risk of health problems like heart diseases, diabetes, high blood pressure and so on

Figure 5: Varying contexts of the word "blood"

5. Discussion

The findings of this study indicate that discussions on health information in Bangladesh heavily emphasize the use of natural ingredients and treatments. Nodes containing natural ingredients such as "juice" exhibit high out-degree and Katz centrality scores. However, the majority of discussions focus on health benefits as a preventive measure rather than a cure, reflecting the broad nature of the general health information domain and its context-dependency. Consistent with Grimes (2021), "hot water" emerged as a top-occurring bigram in the misinformation network, aligning with its propagation as a cure. Similarities with previous studies on health misinformation in India (Naeem, Bhatti and Khan, 2021) were noted, particularly regarding the

prevalence of Ayurvedic practices. However, this study emphasizes the importance of contextual understanding, especially in distinguishing misinformation. Native plants like neem leaves, thankuni leaves, and turmeric, renowned for their medicinal properties in Bengal (Rahmatullah et al., 2010), were prominently featured in the networks. Common diseases in Bangladesh, such as blood pressure and heart diseases, were also key discussion areas in both networks. Despite similarities in lexical usage, the semantic meaning of words differed, posing challenges in misinformation detection without contextual understanding. Clusters formed by the Louvain algorithm revealed misinformation themes related to sexual health, consistent with prior literature (Banik et al., 2023). While one word related to religion was present in the misinformation network, further evidence to elucidate the correlation between religion and misinformation was lacking. Notably, a large cluster centered around the word "hair," aligns with findings on misinformation in dermatology (Iglesias-Puzas et al., 2021). However, the specific relationship between hair-related health information and misinformation in Bangladesh remains unclear. One key limitation of the study is that only one clustering algorithm was used. The study's strength lies in using labeled data validated by health practitioners, enhancing data reliability. Additionally, the comparative analysis between the two networks revealed similar word usage, even in misinformation posts, often employing scientific language for credibility. Focusing on general health information specific to the region provided insights into how cultural and social contexts influence health misinformation. This work has broader implications, potentially aiding in developing region-specific misinformation detection algorithms and informing policymakers on combat strategies, especially in areas like sexual health.

6. Conclusion

In conclusion, the contextuality of co-occurrence network analysis within healthcare is evident, especially in Bangladesh where health misinformation on social media primarily revolves around natural treatments. This study sheds light on the distinctions and overlaps between valid information and misinformation, capturing the social and cultural context of health information dissemination online. It underscores the prevalence of similar semantics and bigram words in both valid information and misinformation. Further research is needed to explore language nuances in these networks, informing policymaking to counter social media misinformation effectively. Additionally, conducting qualitative thematic analysis could validate emerging insights and elucidate misinformation trends in Bangladesh.

References

- Ahmed, I., Hasan, M., Akter, R., Sarkar, B.K., Rahman, M., Sarker, M.S. and Samad, M.A. (2020). Behavioral preventive measures and the use of medicines and herbal products among the public in response to COVID-19 in Bangladesh: A cross-sectional study. *PLoS one*, 15, p.e0243706.
- Ahmed, W., Vidal-Alaball, J. and Vilaseca, Josep M (2022). A social network analysis of Twitter data related to blood clots and vaccines. *International Journal of Environmental Research and Public Health*, 19, p.4584.
- Al-Zaman, Md Sayeed and Noman, (2021). Social media news in crisis? Popularity analysis of the top nine Facebook pages of bangladeshi news media. *Journal of Information Science Theory and Practice*, 9, pp.18–32.
- Ankomah, A., Anyanti, J. and Oladosu, M., 2011. Myths, misinformation, and communication about family planning and contraceptive use in Nigeria. *Open Access Journal of Contraception*, pp.95-105.
- Banik, S., Khan, Jami, H., Sivasubramanian, M., Dhakal, M. and Wilson, E. (2023). Social determinants of sexual health among sexual and gender diverse people in south asia: Lessons learned from india, bangladesh, nepal, and pakistan. Springer, pp.327–352.
- Cabral, L., Monteiro, J.M., José, S., Lincoln and Jorge, P. (2021). FakeWhastApp. BR: NLP and machine learning techniques for misinformation detection in brazilian portuguese WhatsApp messages. pp.63–74.
- Caulfield, T. (2020). *Pseudoscience and COVID-19-we've had enough already*. *Nature (Lond.)*.
- Cavallo, D.N., Chou, W.-Y.S., McQueen, A., Ramirez, A. and Riley, W.T. (2014). *Cancer prevention and control interventions using social media: user-generated approaches*. AACR.
- Chou, W.-Y.S., Oh, A. and Klein, W.M. (2018). Addressing health-related misinformation on social media. *Jama*, 320, pp.2417–2418.
- Evert, S., 2005. The statistics of word cooccurrences: word pairs and collocations.
- Funk, S., Salathé, M. and Jansen, V.A. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7, pp.1247–1256.
- Fudolig, M.I., Alshaabi, T., Arnold, M.V., Danforth, C.M. and Dodds, P.S. (2022). Sentiment and structure in word co-occurrence networks on Twitter. *Applied Network Science*, 7, pp.1–27.
- Garg, M. and Kumar, M. (2018). Identifying influential segments from word co-occurrence networks using AHP. *Cognitive Systems Research*, 47, pp.28–41.
- Grimes, D.R. (2021). Medical disinformation and the unviable nature of COVID-19 conspiracy theories. *PLoS One*, 16, p.e0245900.

- Gurcan, F., Dalveren, Gonca Gokce Menekse, Cagiltay, Nergiz Ercil, Roman, D. and Soylu, A. (2022). Evolution of software testing strategies and trends: Semantic content analysis of software research corpus of the last 40 years. *IEEE Access*, 10, pp.106093–106109.
- Haque, M.I., Chowdhury, A.A., Shahjahan, M. and Harun, (2018). Traditional healing practices in rural Bangladesh: a qualitative investigation. *BMC complementary and alternative medicine*, 18, pp.1–15.
- Hassanzadeh, M., Khodadust, R. and Zandian, F. (2012). Analysis of co-authorship indicators, betweenness centrality and structural holes of the Iranian nanotechnology researchers in science citation index (1991-2011). *Iranian Journal of Information Processing and Management*, 28, pp.223–249.
- Hossain, I. and Haque, A.M. (2023). *A case study on the spread of fake news by social media in the COVID-19 era in bangladesh*.
- Hu, J., Huang, R. and Wang, Y. (2018). Geographical visualization of research collaborations of library science in China. *The Electronic Library*, 36, pp.414–429.
- Iglesias-Puzas, Á., Conde-Taboada, A., Aranegui-Arteaga, B. and López-Bran, E. (2021). ‘Fake news’ in dermatology. Results from an observational, cross-sectional study. *International Journal of Dermatology*, 60, pp.358–362.
- Jafar, T.H., Gandhi, M., Jehan, I., Naheed, A., Asita, S., H, Shahab, H., Alam, D., Luke, N., Lim, W. and COBRA-BPS Study Group (2018). Determinants of uncontrolled hypertension in rural communities in South Asia—Bangladesh, Pakistan, and Sri Lanka. *American journal of hypertension*, 31, pp.1205–1214.
- Kanozia, R. and Arya, R. (2021). ‘Fake news’, religion, and COVID-19 vaccine hesitancy in India, Pakistan, and Bangladesh. *Media Asia*, 48, pp.313–321.
- Mackey, T.K., Purushothaman, V., Haupt, M., Nali, M.C. and Li, J. (2021). Application of unsupervised machine learning to identify and characterise hydroxychloroquine misinformation on Twitter. *The Lancet Digital Health*, 3, pp.e72–e75.
- Massey, P.M., Kearney, M.D., Hauer, M.K., Selvan, P., Koku, E. and Leader, A.E. (2020). Dimensions of misinformation about the HPV vaccine on Instagram: Content and network analysis of social media characteristics. *Journal of medical Internet research*, 22, p.e21451.
- Mostafa, Z.M., Choudhury, S.R., Ahmed, J., Talukder, M.H. and Rahman, A.S. (2016). Blood glucose and cholesterol levels in adult population of Bangladesh: results from STEPS 2006 survey. *Indian Heart Journal*, 68, pp.52–56.
- Naeem, S.B., Bhatti, R. and Khan, A. (2021). An exploration of how fake news is taking over social media and putting public health at risk. *Health Information & Libraries Journal*, 38, pp.143–149.
- Newman, M. (2018). *Networks*. Oxford university press.
- Nistor, A. and Zadobrischi, E. (2022). The influence of fake news on social media: analysis and verification of web content during the COVID-19 pandemic by advanced machine learning methods and natural language processing. *Sustainability*, 14, p.10466.
- Pulido, C.M., Ruiz-Eugenio, L., Redondo-Sama, G. and Villarejo-Carballido, B., 2020. A new application of social impact in social media for overcoming fake news in health. *International journal of environmental research and public health*, 17(7), p.2430.
- Rahmatullah, M., Khatun, M.A., Morshed, N., Neogi, Prashanta Kumar, Ahmed, Hossan, Md S, Mahal, M.J. and Jahan, R. (2010). A randomized survey of medicinal plants used by folk medicinal healers of Sylhet Division, Bangladesh. *Advances in Natural and Applied Sciences*, 4, pp.52–62.
- Rozario, S. (2009). Allah is the scientist of the scientists: Modern medicine and religious healing among British Bangladeshis. *Culture and religion*, 10, pp.177–199.
- Safarnejad, L., Xu, Q., Ge, Y., Krishnan, S., Bagarvathi, A. and Chen, S. (2020). Contrasting misinformation and real-information dissemination network structures on social media during a health emergency. *American journal of public health*, 110, pp.S340–S347.
- Sentell, T., Kearney, M., Hazen, A., Dela Cruz, M.R., Yamauchi, J., McHenry, L., Rodericks, B. and Massey, P., 2020. Do you Vape With Aloha?: How culture and place are used to spread misinformation on Instagram. *European Journal of Public Health*, 30(Supplement_5), pp.ckaa165-1005.
- Simon, K. (2021). *Digital 2022: Bangladesh. Datareportal* <https://rb.gy/dl37o>.
- Suryadi, D. and Kim, H. (2019). Automatic identification of product usage contexts from online customer reviews. In: *Cambridge University Press*. pp.2507–2516.
- Trew, B. (2020). *Coronavirus: hundreds dead in Iran from drinking methanol amid fake reports it cures disease. Independent*.
- Voitalov, I., Der, V., Der, V. and Krioukov, D. (2019). Scale-free networks well done. *Physical Review Research*, 1, p.033034.
- Vosoughi, S., Roy, D. and Aral, S. (2018). The spread of true and false news online. *science*, 359, pp.1146–1151.
- Wang, Y. (2018). Systematic review on the social mechanism of health misinformation dissemination in the Internet era. *European Journal of Public Health*, 28, pp.cky213-194.
- Xiong, F. and Liu, Y. (2014). Opinion formation on social media: an empirical approach. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24.