A Machine Learning Application for Raising WASH Awareness in the Times of COVID-19 Pandemic

Rohan Pandey  
Shiv Nadar University

Vaibhav Gautam  
Shiv Nadar University

Ridam Pal  
Indraprastha Institute of Information Technology Delhi

Harsh Bandhey  
Indraprastha Institute of Information Technology Delhi

Lovedeep Singh Dhingra  
All India Institute of Medical Sciences

Vihaan Misra  
Netaji Subhas University of Technology

Himanshu Sharma  
GL Bajaj Institute Of Tech and Management

Chirag Jain  
Indraprastha Institute of Information Technology Delhi

Kanav Bhagat  
Indraprastha Institute of Information Technology Delhi

Arushi Arushi  
All India Institute of Medical Sciences

Lajjaben Patel  
All India Institute of Medical Sciences

Mudit Agarwal  
All India Institute of Medical Sciences

Samprati Agrawal  
All India Institute of Medical Sciences

Rishabh Jalan  
Indraprastha Institute of Information Technology Delhi

Akshat Wadhwa  
Indraprastha Institute of Information Technology Delhi

Ayush Garg  
Indraprastha Institute of Information Technology Delhi

Yashwin Agrawal
Research Article

Keywords: COVID-19, mHealth, Machine Learning, chatbot, Natural Language Processing.

Posted Date: June 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-562183/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Scientific Reports on January 17th, 2022. See the published version at https://doi.org/10.1038/s41598-021-03869-6.
Abstract

Background: The COVID-19 pandemic has uncovered the potential of digital misinformation in shaping the health of nations. The deluge of unverified information that spreads faster than the epidemic itself is an unprecedented phenomenon that has put millions of lives in danger. Mitigating this 'Infodemic' requires robust health messaging systems that are engaging, vernacular, scalable, effective, and continuously learn new misinformation patterns.

Objective: We created WashKaro, a multi-pronged intervention for mitigating misinformation through conversational Artificial Intelligence (AI), machine translation and natural language processing (NLP). WashKaro provides the correct information matched against WHO guidelines through AI and delivers it in a suitable format in local languages.

Results: A total of 5026 people downloaded the app during the study window; among those, 1545 were actively engaged users. Our study shows that 3.4 times more females engaged with the App in Hindi as compared to males, the relevance of AI-filtered news content doubled within 45 days of continuous machine learning, and the prudence of integrated AI chatbot “Satya” increased thus proving the usefulness of a mHealth platform to mitigate health misinformation.

Conclusion: We conclude that a machine learning application delivering bite-sized vernacular audios and conversational AI is a practical approach to mitigate health misinformation.

1 Introduction

Healthcare misinformation is a growing menace in digital societies[1, 2]. This is clearly highlighted by the COVID-19 pandemic that has affected over 3.8 million people worldwide, causing a widespread loss in all aspects of daily life [3]. Digital consumption has increased manifolds, creating both an opportunity and a danger in terms of information dissemination. Infodemic has been defined as an overabundance of information, some accurate and some not, making it hard for people to find trustworthy sources and reliable guidance when they need it [4]. The spread of the COVID-19 infodemic was much faster than the pandemic itself and poses a threat to public health [5]. Further, mitigation of misinformation is also vital for raising correct awareness for the primary prevention of most communicable and non-communicable diseases. Mobile health (mHealth), coupled with verified health information, can serve as an information dispensing tool to tackle the spread of misinformation. Clear and effective communication of preventive measures and updated information is essential. To achieve this goal, designing a trustworthy app that helps navigate the information deluge can be crucial. Therefore, recognizing the potential of mHealth platforms, we developed WashKaro, a multi-pronged AI approach for Infodemic Management. WashKaro was driven by the imminent need to raise Water, Sanitation, and Hygiene (WASH) awareness and combines English (WASH) with vernacular (Karo, meaning "Do" in Hindi) for mitigating the spread of COVID-19. OnAir is a feature on the WashKaro app which combines Natural Language Processing (NLP) to match news articles with WHO guidelines. Conversational AI (Satya, meaning "Truth" in Hindi) reaches
out to the community as audio-visual content in local languages. To keep the information relevant, WashKaro provides daily news matched with WHO guidelines [6], WHO directive-based Symptom Self-Assessment tool, and human-vetted information delivering these in Hindi, the most widely understood local language across India. Since India is one of the largest and fastest-growing markets for digital consumers, with 560 million Internet subscribers in 2018 [7], and about 60% using mHealth technologies [8], this offered a unique opportunity to test WashKaro. The study is based on the WHO’s Information Network for Epidemics (EPI-WIN) [9] strategy, covering four strategic areas of work to tackle the infodemic as shown in Fig. 1.

The methodology used for this study is centered around the WHO EPI-WIN strategy, which covers four strategic areas of work to respond to the infodemic. The first area of coverage is focused on identifying the problem at hand, given the current evidence and information to promote and form public policies strategically. We have identified the context-specific community problems and the potential of mHealth, AI, and NLP in order to acknowledge possible solutions. This is followed by simplifying the enormous amount of information currently available across multiple sources to disseminate accurate information in a simplified manner. In order to achieve this objective, multilingual support is provided in the form of audio visual-based content. The spread of misinformation is tackled by providing information such as Mythbusters and government updates along with meta-information in the form of geographic coordinates of essential facilities. We also offer periodic hand washing reminders. In case of such an unforeseen event, it is vital to amplify the intervention by means of establishing two-way communication with the intended audience to tailor the advice and messages. This has been catered to by engaging users in active feedback based involvement, participating in enhancing the AI proposed model along with any generic feedback in an audio format. A public health survey and symptom self-assessment are crucial components in amplifying our study. In order to devise constantly evolving strategies, it is essential to validate the methodology and quantify the infodemic. WashKaro application statistics, demographic analysis of public health surveys, health analysis of at-risk population using symptom self-assessment, and user agreement on AI-based intervention is critical to quantify and evaluate.

2 Methods

WashKaro was developed as a holistic mHealth solution that could serve as a one-stop AI-powered infodemic management suite during the current COVID 19 pandemic. The underlying strategy utilized was Identify-Simplify-Amplify-Quantify, as deployed by the Information Network for Epidemics (EPI-WIN) established by the WHO [4]. The main idea of our application was to provide unsolicited information as bite-sized text and audio in Hindi and English. The machine learning algorithms helped in filtering correct information from the deluge of news, which were then vetted by medical experts. Gathering raw data from credible sources such as the WHO and consumer-centric daily news articles, we used NLP approaches and Machine Learning (ML) to identify authentic and pertinent information. The information thus extracted was simplified and presented as audio-visual content in Hindi (the most widely understood local language across India), English, and various other vernaculars. By garnering feedback on the
relevance of the WHO information provided along with the news pieces, the advice to the individual was tailored according to their personal needs, thus amplifying the reach of appropriate messages. We also offered a WHO directive-based Symptom Self-Assessment tool and numerous categories of human-vetted information in the form of Infographics, MythBusters, geographic information, etc. Forming real-time, on-the-ground, multidisciplinary research partnerships is essential to mitigate the infodemic. Our entire methodology and infodemic suite is, therefore, open-source (https://github.com/tavlab-iiitd/WashKaro/tree/master/washkaro-textmatching) and available for the whole of the scientific community to build upon it.

2.1 NLP in healthcare

In the current situation, timely delivery of tenable content to the masses is exceptionally crucial to counter the spread of misinformation. WashKaro targeted this requirement using Natural Language Processing techniques to dispense information sourced through highly trusted outlets of WHO such as EPI-WIN, which perhaps may not reach the appropriate audience, or be too complicated for them [10].

The NLP pipeline (Fig. 2) involved two datasets: the WHO guidelines and the news articles. Multiple pairs of WHO guidelines and News Articles are generated as an input for the Machine Learning System, extractive ML summarization techniques were used to abbreviate the text. Articles were refreshed on a daily basis (using an automated web scraper) and from the Indian vernacular news source: Dainik Jagran. Pre-Trained Word2Vec Embedding [11] was used to generate embedding vectors for each word in the two documents. These word vectors were converted to article-level embeddings using Smooth Inverse Frequency [12]. The generated pair of document level vectors are used for the calculation of distance metrics. Cosine similarity was calculated to find the similarity between two embedded vectors. Based on the users' reviews, the threshold of cosine similarity was set to determine the news articles that will be provided to users subsequently using this AI system[13]. This pipeline served to complement the user's daily news consumption that suits their palette with an appropriate WHO guideline related to COVID-19 and WASH (Water Sanitation and Hygiene), thus augmenting healthcare awareness. In order to enhance engagement and provide increasingly relevant content, user feedback was sought at the end of each matching- the users marked each pair of WHO guideline and news article provided to them as either relevant or irrelevant. This active user feedback aided the machine learning backend in improving with each review by determining the type of news articles the user found relevant to a particular guideline. Further, any new article provided to the user took into account the previous learning, which enabled the deliverance of more relevant information with each feedback cycle.

2.2 Simplification

We made a deliberate attempt to convey context-specific and consumable information in a medium of the user's choice. Infographics based on WHO recommendations were used for effective presentation of preventive measures. Byte sized information packets were delivered in multiple local languages to ensure accessibility to various marginalized groups. Text-to-speech engines helped convert the information to an audio-visual format, thus reaching out to the less educated population. The application made use of the
inbuilt Android text-to-speech model available in each android device. Mythbusters and government
advisories, critical in countering misinformation and uncertainties surrounding the official guidelines,
were credibly sourced and regularly updated. Mythbusters and government advisories, critical in
countering misinformation and uncertainties surrounding the official guidelines, were credibly sourced
from the official government website of the Ministry of Health and Family Welfare [14]. The guidelines
related to COVID-19 present in the MOHFW were updated on a daily basis, and the WashKaro application
was simultaneously updated. Information on containment zones, hospitals, and hunger relief centers
was provided in a geographical context, with directions imparted through Google Maps, a popular user-
friendly interface. Regular notifications, worded positively to encourage participation, reminding the user
to wash their hands and use masks in public places, were displayed.

2.3 Symptom Self Assessment

Low accessibility of the healthcare system, given the lockdown and social distancing measures in place,
and a skewed ratio between the population who wishes to get tested and medical professionals who can
verify this need call for an effective alternative to screen patients [15]. Thus, we devised a self-
assessment tool for the symptoms of COVID-19, thereby enabling quicker identification of suspect cases
who can then be guided to the Government helpline numbers and informed about proper self-quarantine
protocols, nearby hospitals admitting COVID suspects, and testing centers. We defined the Suspect Case
using the WHO Interim Guidance on Global surveillance for COVID-19 caused by human infection with the
COVID-19 virus, and classified them further as a Suspect case (A), (B) or (C) [16]. The 7-point
questionnaire was designed using the case definitions from the WHO Interim Guidance verbatim. Based
on the WHO criteria' application on the answers to the 7 questions, the user was notified about whether or
not they were suspected of having COVID-19 (Fig. 3).

2.4 Chatbot

Correct and officially verified information regarding the disease should be there for everyone's disposal.
To serve this purpose, we have made a chatbot (Fig. 4), which has verified information from WHO, CDC,
and additional government-approved sources. Existing Solution consists of an option Driven System[17]
where a user needs to select through various lists of options to find answers to the Query. Thus, we
devised a chatbot system designed to answer user queries using natural language. The current system
consists of a Long Short Term Memory (LSTM) model fine-tuned on a Medical Question-Answers Dataset
(MedQuAD) dataset [18]. The Dataset was encoded using swivel embeddings generated on the Covid-19
open research dataset[19]. All the data, including the training set, is incubated from Credible and
Government controlled sources. Data for Answering the input query was taken from three Sources. Daily
Statewise and India case count were extracted and updated on a real-time basis[20]. Data for Training the
model was taken from the World Health Organisation's CoronaVirus Frequently Asked Questions[21] and
Centers for Disease Control and Prevention's CoronaVirus Frequently Asked Questions[22]. The user Query
is also passed through spelling correction using symmetric delete spelling correction algorithm along
with an artificial increasing frequency of words related to the disease, symptoms, etc.[23] so as to increase the accuracy and effectiveness of the system.

2.5 Active User Feedback

Anonymized data was collected through self-assessment usage analytics, Play Store managed user statistics, and easy to comprehend survey forms. Our suite deployed an anonymized public health survey that asked basic healthcare-related questions to understand the demographics and monitor the situation periodically. To ensure user convenience, establish a two-way dialogue and prevent specific suggestions from being marginalized, audio-based feedback was taken from the user. Illustrations were used whenever possible to make the user aware of the data being collected, hence protecting their right to information and privacy.

3 Results

The study aimed to improve article relevance using a machine learning approach and achieve information dissemination through better engagement in the local language. Our multi-pronged approach targeted to:

(i) Achieve a non-intrusive manner of healthcare knowledge dissemination.

(ii) Use local language to increase the participation of the target group (Female population).

(ii) Develop a self-assessment tool to identify the at-risk population at an early stage to mitigate the chances of community transmission.

3.1 Information Enrichment Over Time: The Number of `Relevant' Votes Increases

With time and increased user feedback, the relevance of the matching of news articles with WHO reports increases as seen in the rise in the number of relevant votes, owing to the constantly evolving machine learning model (Fig. 5). The number of irrelevant votes also decreases, validating our proposed methodology and providing increasingly relevant content from trusted sources to the user over time in the language of their preferred choice. At the beginning of the AI-based learning system on 15 March 2020, the number of relevant votes and irrelevant votes were both 18. On 25 March 2020, with an increase in user interaction and AI learning, the number of relevant votes was 173 and irrelevant votes were 69. The ratio of relevant votes to irrelevant votes increased from 1.0 to 2.5 over a period of one month.

3.2 Demographics- Females engaged more in Hindi

A total of 436 people took part in the English language based survey, and 126 took part in the Hindi language based survey. Fig. 6 was generated based on the survey conducted on the WashKaro app. The analysis of this plot suggests that the number of English users are more than the number of Hindi users. It also depicts that the overall number of male users is more than female users. A key insight observed
from the data depicted that Hindi speaking female users (33% of total Hindi speaking users) were more than English speaking female users (11% of total English speaking users). The census of India 2011, highlighted the disparity of literacy rates across genders with 82.14% literacy rates amongst Indian males and 65.46% literacy rates in Indian females [24]. This underscores the fact that using local languages empowers the sections of the population that might not have otherwise access to the information.

3.3 Target population: Users who reported higher than expected incidence

Based on the data collected from the symptom self-assessment, we can distil the general population so that we can trace the potential number of positive patients at risk of COVID-19 and help identify patients who are already suffering from this disease. The question has been framed in such a way that potential threats are identified at the early stage so that we can mitigate the chances of community transmission. Out of 3567 respondents, 276 (7% of respondents) were found to be suspect cases according to their responses, while 3291 were non-suspect. 467 (13.09%) of the respondents reported a travel history to locations reporting community transmission of COVID-19, 326 (9.13%) reported close contact with COVID-19 positive patient, 323 (9.05% of respondents) reported fever, 556 (15.58%) reported cough, 367 (10.28%) reported shortness of breath (SOB), 277 (7.77%) reported that they required Hospitalization and 395 (11.07%) respondents reported there was an alternate diagnosis for their condition (Fig. 7).

4 Discussion

Over the years mHealth and machine learning have made significant contributions in the medical domain[25]. In the case of COVID-19, where reliable therapeutic strategies are still under experimentation, the role of such mHealth and machine learning interventions is critical [26, 27, 28]. Timely dissemination of trusted byte-sized information is extremely instrumental in mitigating the infodemic. With the majority of the world population staying at home due to lockdown, the increased amount of digital consumption opens up the scope to deploy these techniques as an effective social intervention to mitigate the infodemic by delivering the right information to the right people at the right time. When organizations across the globe are proactively testing various strategies to address the issue at hand, open-source software will play a vital role in such scenarios at a global stage to mitigate pandemics and infodemics at the root level. We provide an open-source template featured with functions like Symptom Self-Assessment, Notification amplifier which notifies the user for washing hands, which are required for fighting against epidemics and pandemics. This also helps in the propagation of the right information hence proper management of infodemic can be done without misleading the masses in crucial scenarios. The enhancement of a few features within the WashKaro application can help in serving as an effective intervention for the government and policymakers. Detailed questions can be formulated targeting the at-risk users identified using symptom self-assessment, which can be incorporated into the existing framework followed by the higher authorities and medical workers for predicting the suspects of COVID-19 at an early stage. After identification and testing of the at-risk population, our analysis can be extended to predict patients who have chances of being at risk in the near future from the definitive set of questions, based on the priority of each question. We can present this data to the suitable administration
and decision-makers for taking effective measures against such individuals at an early stage. For the Suspect Cases, we can administer a second questionnaire to further stratify the risk of acute respiratory distress syndrome (ARDS) and septic shock by assessing the severity of symptoms and looking for identified risk factors like age and pre-existing comorbidities that are not included in the WHO Interim Guidance. This can aid in making decisions regarding home quarantine against hospital admission. The app can also be used to identify other suspect cases in the same household. Further, to assist the government authorities to identify those requiring testing for COVID-19, we can ask for contact details of the Suspect Cases with informed consent and relay them to the appropriate government authorities to enable targeted testing. A follow up of the suspect cases through push notifications, advising testing and recording test results, can help ensure that complacency does not set in. Also, as highlighted by previous work, teaching interventions to women can be an important step for mitigating diseases [29]. We reason that the engagement in Hindi is due to the content rather than the nature of the App because the default language of the app content is English. Therefore, the users would have made a conscious effort to change the language to Hindi.

Finally, there are some limitations to the study that have been conducted prior to the revamp of the application. All applications with COVID-19 information were removed in association with the guidelines regarding COVID-19 related applications on the Google Play store [30]. The time frame of the case study for COVID-19 was shortened due to this reason. Our application was the first application providing vetted information using machine learning. As the official apps were rolled out by the government, these were promoted on a larger scale to target the national audience. That may explain the drop in the number of users active on our app. Secondly, due to data privacy policies, we could not obtain granular information on individuals' locations, such granularity of data could have helped articulate user studies in a more diverse way. The development of innovative approaches while protecting individual data yet gathering useful inference is an active area of research, and our further work will address this limitation in various public health scenarios. Therefore, we conclude that the role of digital health interventions in the form of systems articulating vetted messages needs to be explored effectively dealing with public health challenges, both during health emergencies and normal times addressing the Sustainable Development Goals (SDGs) put forward by WHO.

Since health information from credible sources is not necessarily prioritized for dissemination in conventional media, especially for less literate and non-English speaking sections of the society. The users were searching for such information at a time when little was available in a local language in India. Our App was the first such platform and >5000 people downloaded the app during the study window, among those 1545 were actively engaged users before Google pulled the plug on all COVID-19 apps. Going forward, we are continuing to develop natural language processing based services for extending this feasibility experiment for raising public health awareness about infectious diseases such as TB. The further scope and extension of this study involve quizzes, gamification, peer network building and social incentivization to engage users. Future work is planned to evaluate the human-centric design of the app collaboration with an NGO working on the ground. Our use of models to manage information through the use of machine learning optimizes human resources and is shown to be effective as per the measured
parameters. We strongly believe that this approach may be relevant to other resource-constrained settings also.

**Abbreviations**

WHO: World Health Organization

WASH: Water Sanitation Hygiene

EPI-WIN: WHO’s Information Network for Epidemics

AI: Artificial Intelligence

NLP: Natural Language Processing

**Declarations**

**Author Contributions**

RP - Language models, OnAIr Engine, Paper writing, team coordination

VG - App development, backend

RP - NLP and AI model development and validation, paper writing, Review

HB - Writing, Editing, App Development

LSD - ISAQ oriented design for App. Coordination with clinical teams, workflow and design of Symptom Checker

VM - App development, app maintenance on Google play store, writing, editing

HS - App development, backend

CJ - App development, schemas for the Firebase backend

KB - Data Collection, Similarity matching techniques, Testing the algorithms

A - Design of clinical algorithms in the app, paper writing.

PLJ - Design of clinical algorithms in the app, paper writing.

MA - Design of clinical algorithms in the app, paper writing.

SA - Design of clinical algorithms in the app, paper writing.
RJ - Satya Chatbot engine
AW - Satya Chatbot data sources and integration
AG - Satya Chatbot backend
YA - App development, UI design
BR - UI Design, Information Curation
PK - Co-investigator, app design and team building
TS - Principal investigator, conceptualized the study, study design, team building, paper writing.

**Funding**

This work was partly supported by the Wellcome Trust/DBT India Alliance Fellowship IA/CPHE/14/1/501504 awarded to Tavpritesh Sethi.

**Competing Interests**

The authors declare no competing interests.

**Acknowledgments**

Tavpritesh Sethi also acknowledges the Delhi Knowledge Development Foundation supported Center of Excellence in Healthcare at IIIT-Delhi and The Wellcome Trust DBT India Alliance IA/CPHE/14/1/501504 for salary support. We also acknowledge Prof. Rakesh Lodha, AIIMS New Delhi, for his clinical inputs.

**References**

1. Ioannidis, John PA, et al. "How to survive the medical misinformation mess." European journal of clinical investigation 47.11 (2017): 795-802. doi: 10.1111/eci.12834
2. Tasnim, Samia, Md Mahbub Hossain, and Hoimonty Mazumder. Impact of rumors and misinformation on COVID-19 in social media. Journal of preventive medicine and public health 53.3 (2020): 171-174. doi: 10.3961/jpmph.20.094
3. World Health Organization, et al.. Coronavirus disease 2019 (COVID-19): situation report, 110; 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200509covid-19-sitrep-110.pdf
4. World Health Organization, et al.. Coronavirus disease 2019 (COVID-19): situation report, 91; 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200420-sitrep-91-covid-19.pdf
5. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: Quantifying the covid-19 misinformation epidemic on twitter. Cureus. 2020;12(3). doi:10.7759/cureus.7255

6. Novel Coronavirus Technical Guidance, World Health Organization. URL: www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance

7. Kaka N, Madgavkar A, Kshirsagar A, Gupta R, Manyika J, Bahl K, et al. Digital India: Technology to transform a connected nation. McKinsey Global Institute, March. 2019. URL: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/digital-india-technology-to-transform-a-connected-nation

8. Levy D, Wasden C, DiFilippo D, Sur P. Emerging mHealth: paths for growth. PwC M-Health. 2012; p. 1-44. URL: https://www.pwc.com/gx/en/healthcare/mhealth/assets/pwc-emerging-mhealth-full.pdf

9. World Health Organization's Information Network for Epidemics, EPI-WIN Updates. World Health Organization. URL: https://www.who.int/teams/risk-communication/epi-win-updates

10. Zarocostas J. How to fight an infodemic. The Lancet. 2020;395(10225):676. doi:10.1016/S0140-6736(20)30461-X

11. Mikolov T, et al. Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems. 2013. ArXiv. doi: arXiv:1310.4546v1

12. Arora S, Liang Y, and Ma T. A simple but tough-to-beat baseline for sentence embeddings. Proceedings of International Conference on Learning Representations. 2017. URL: https://openreview.net/forum?id=SyK00v5xx

13. Pal R, et al. A Cross-lingual Natural Language Processing Framework for Infodemic Management. ArXiv;2020. doi: arXiv:2010.16357

14. Ministry of Health and Family Welfare. Ministry of Health and Family Welfare Government of India. URL: https://www.mohfw.gov.in/

15. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al.. Fair allocation of scarce medical resources in the time of Covid-19; 2020. doi:10.1056/NEJMsb2005114

16. World Health Organization, et al. Global surveillance for COVID-19 caused by human infection with COVID-19 virus: interim guidance, 20 March 2020. World Health Organization; 2020. URL:https://apps.who.int/iris/handle/10665/331506

17. WHO Health Alert. 20 March 2020. World Health Organization; 2020. URL:https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp [Accessed October 7th, 2020]

18. Abacha AB, Demner-Fushman D. A Question-Entailment Approach to Question Answering. ArXiv; 2019. doi:10.1186/s12859-019-3119-4

19. Wang, Lucy Lu et al. “CORD-19: The Covid-19 Open Research Dataset.” ArXiv; 2020. doi: arXiv:2004.10706.
21. MyGov.in covid19 state wise status. URL:https://www.mygov.in/corona-data/covid19-statewise-status/
22. Q&As on Covid and related health topics. World Health Organisation. URL:https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19
23. Coronavirus (COVID-19) Frequently Asked Questions. Centers for Disease Control and Prevention. URL:https://www.cdc.gov/coronavirus/2019-ncov/faq.html
24. Coronavirus (COVID-19) Outbreak Glossary. Kaiser Family Foundation; 2020. URL:https://www.kff.org/glossary/covid-19-outbreak-glossary/
25. Chandramouli C, General R. Census of India 2011. Provisional Population Totals New Delhi: Government of India. 2011; p. 409-413. URL:https://censusindia.gov.in/2011-prov-results/paper2/data_files/india/paper2_1.pdf
26. Stepanian, Robert SH, and Turki Al-Anzi. m-Health 2.0: new perspectives on mobile health, machine learning and big data analytics. Methods 151 (2018): 34-40. doi: 10.1016/j.ymeth.2018.05.015
27. Alimadadi, Ahmad, et al. Artificial intelligence and machine learning to fight COVID-19. Physiological Genomics, 52(4), 200–202. (2020). doi:10.1152/physiolgenomics.00029.2020
28. Wang, Peipei, et al. Prediction of epidemic trends in COVID-19 with logistic model and machine learning techniques. Chaos, Solitons & Fractals 139 (2020): 110058. doi: 10.1016/j.chaos.2020.110058
29. Lalmuanawma, Samuel, Jamal H, and Lalrinfela C. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos, Solitons & Fractals (2020): 110059. doi: 10.1016/j.chaos.2020.110059
30. Caldwell JC, Caldwell P. Roles of women, families and communities in preventing illness and providing health services in developing countries. The epidemiologic transition: policy and planning implications for developing countries Washington DC: National Academy Press p. 1993; p. 252-271. URL: https://www.nap.edu/read/2225/chapter/14
31. Requirements for coronavirus disease 2019 (COVID-19) apps, PlayStore, Google. URL:https://support.google.com/googleplay/android-developer/answer/9889712?hl=en

Figures
Figure 1

Proposed workflow of the App based upon Identify, Simplify, Amplify and Quantify framework as specified WHO’s EPI-WIN strategy [4].
Figure 2

NLP Pipeline. The pipeline takes in news articles and the World Health Organization (WHO) reports and constructs two-level sentence similarity between titles and the full-text to build a similarity score. Finally, the relevant texts are subject to translation and text to speech conversion for local language consumption (Hindi).
Figure 3

Self Assessment Tool Flowchart. Based on the World Health Organization (WHO) Interim Guidance, a questionnaire and flowchart were developed to classify the responders as *Suspects* or *Non-suspects*. Here SOB refers to Shortness Of Breath, and ARI refers to Acute Respiratory Infection.
Figure 4

Request-Response cycle in the chatbot. This is a schematic diagram depicting how the answer is displayed whenever a query is asked to the chatbot by a user.

Figure 5
Analysis of Natural Language Processing (NLP) pipeline. The graph shows relevance as a function of user feedback functionality in the app. Relevance is seen to increase with cumulative feedback over time. From day 0 onwards, the Relevant Count’s angular coefficient is 1.39 (±0.488), angular coefficient of Irrelevant Count is -0.99 (±0.602), with an average slope difference of about 2.29.

![Graph showing relevance over time](image)

**Figure 6**

Analysis of Public Health Survey. Distribution graphs showing the distribution of gender among Hindi and English Users. It clearly shows skewness in gender for English users whereas in the case of Hindi users it shows an approximate normalization among the genders.

![Bar chart showing gender distribution](image)
Figure 7

Analysis of Self Assessment. A simple user-level self-assessment has been deployed to enable the general population to perform self-assessment and identify the population at risk, which can be used as an effective screening. This vividly states that the person tested for self-analysis was higher than expected numbers.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.pdf