

Natural language processing to evaluate texting conversations between patients and healthcare providers during COVID-19 Home-Based Care in Rwanda at scale

Richard T Lester^{1¶*}, Matthew Manson^{1¶}, Muhammed Semakula^{2,3}, Hyeju Jang^{4,5}
Hassan Mugabo³, Ali Magzari⁶, Junhong Ma Blackmer⁷, Fanan Fattah¹, Simon Pierre
Niyonsenga³, Edson Rwagasore³, Charles Ruranga⁸, Eric Remera³, Jean Claude S.
Ngabonziza^{3,9}, Giuseppe Carenini⁵, Sabin Nsanzimana²

¹Division of Infectious Diseases, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

²Rwanda Ministry of Health, Kigali, Rwanda

³Rwanda Biomedical Centre, Kigali, Rwanda

⁴Luddy School of Informatics, Computing, and Engineering, Department of Computer Science Indiana University Indianapolis, Indianapolis, Indiana, United States

⁵Department of Computer Science, Faculty of Science, University of British Columbia, Vancouver, British Columbia, Canada

⁶Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, British Columbia, Canada

⁷Department of Mathematics, University of British Columbia, Vancouver, British Columbia, Canada

⁸African Center of Excellence in Data Science, University of Rwanda, Kigali, Rwanda

⁹Department of Clinical Biology, University of Rwanda, Kigali, Rwanda

*Corresponding author

Email: [rl Ester@mail.ubc.ca](mailto:rl Lester@mail.ubc.ca)

¶ These authors contributed equally to this work.

Abstract

Isolation of patients with communicable infectious diseases limits spread of pathogens but can be difficult to manage outside hospitals. Rwanda deployed a digital health service nationally to assist public health clinicians to remotely monitor and support SARS-CoV-2 cases via their mobile phones using daily interactive short message service (SMS) check-ins. We aimed to assess the texting patterns and communicated topics to understand patient experiences. We extracted data on all COVID-19 cases and exposed contacts who were enrolled in the WelTel text messaging program between March 18, 2020, and March 31, 2022, and linked demographic and clinical data from the national COVID-19 registry. A sample of the text conversation corpus was English-translated and labeled with topics of interest defined by medical experts. Multiple natural language processing (NLP) topic classification models were trained and compared using F1 scores. Best performing models were applied to classify unlabeled conversations. Total 33,081 isolated patients (mean age 33.9, range 0-100), 44% female, including 30,398 cases and 2,683 contacts) were registered in WelTel. Registered patients generated 12,119 interactive text conversations in Kinyarwanda (n=8,183, 67%), English (n=3,069, 25%) and other languages. Sufficiently trained large language models (LLMs) were unavailable for Kinyarwanda. Traditional machine learning (ML) models outperformed fine-tuned transformer architecture language models on the native untranslated language corpus, however, the reverse was observed of models trained on English-only data. The most frequently identified topics discussed included symptoms (69%), diagnostics (38%), social issues (19%),

prevention (18%), healthcare logistics (16%), and treatment (8.5%). Education, advice, and triage on these topics were provided to patients. Interactive text messaging can be used to remotely support isolated patients in pandemics at scale. NLP can help evaluate the medical and social factors that affect isolated patients which could ultimately inform precision public health responses to future pandemics.

Author Summary

We present the first application of NLP for categorizing text messages between patients and healthcare providers within a nationally scaled digital healthcare program. This study provides unique insights into the circumstances of home-based COVID-19 patients during the pandemic. Our trained topic classification models accurately categorized topics in both English and African language texts. Patients reported and discussed both medical and social issues with public healthcare providers. This approach has the potential to guide precision public health decisions and responses in future outbreaks, pandemics, and remote healthcare scenarios.

Introduction

The practice of isolating patients with communicable infectious diseases reduces transmission by separating infected individuals from healthy ones, and isolation at home allows healthcare systems to focus on managing severe cases in hospitals.¹ However, isolated patients may have changes to their condition that require attention from medical

professionals and can experience additional social stressors.² Isolation protocols have even been considered a risk factor for mortality.³ Understanding the complexity of facility and home-based isolation remains a critical consideration in effective pandemic response planning.

At the pandemic outset, the Rwanda government COVID-19 Joint Task Force adopted a short message service (SMS)-based digital health service, WeTel, to remotely monitor and support community isolated cases and contacts of SARS-CoV-2 from the centralized national command post. Its use continued throughout the emergency phase of the pandemic as the key to their Home-Based Care (HBC) program.⁴ The WeTel platform was selected based on its history of use in HIV programs, prior clinic evidence, and importantly, its accessibility.⁵⁻⁷ Internet based tools and smartphone apps would have been insufficient since, as a lower-middle income country (LMIC), only 23% of the Rwanda population has internet-access; yet at least 90% of the population has access to cellular phones.^{8,9} As in other applications of the WeTel platform, throughout the COVID-19 response in Rwanda, automated 'check-in' messages were sent daily to isolated patients, to which they could respond to report on their status and ask questions. A manually performed topic classification of WeTel text messaging conversations with patients during HIV care previously demonstrated that these open-ended text messaging conversations contain wide ranging issues experienced by patients that could be acted upon.¹⁰

Advances in natural language processing (NLP) enables rapid computational topic analysis (classification and modelling) which can delineate large amounts of text into topics of interest.¹¹ Topic classification models using machine-learning (ML) have been used in clinical settings to label text messages exchanged between healthcare providers and outpatients, but never to our knowledge in a national scale program.^{12–15} Unfortunately, the uneven distribution of global healthcare data has the potential to leave entire continents behind in the development of artificial intelligence (AI) for healthcare.¹⁶

In this study, we aimed to analyze the complete corpus of SMS conversations between Rwanda’s public health clinicians and isolated COVID-19 patients and contacts during the first two years of the COVID-19 pandemic. We considered a clinical, social, and logistical perspective using topic classification to better delineate the needs and experiences of patients and the advice they were given, during isolation – a large-scale approach that could be used to guide decision-making in outbreak management.

Results

Participation in text messaging conversations

During the study period, Rwanda reported 131,190 cases of COVID-19, of whom 33,081 individuals (25%, 30,398 cases, 2,683 contacts) were registered in the WelTel

texting service. The numbers of patients registered, and the occurrence of text-message conversations (with at least 3 interactive messages), appeared in waves similar to the COVID-19 incidence reported in the national registry (Fig. 1).

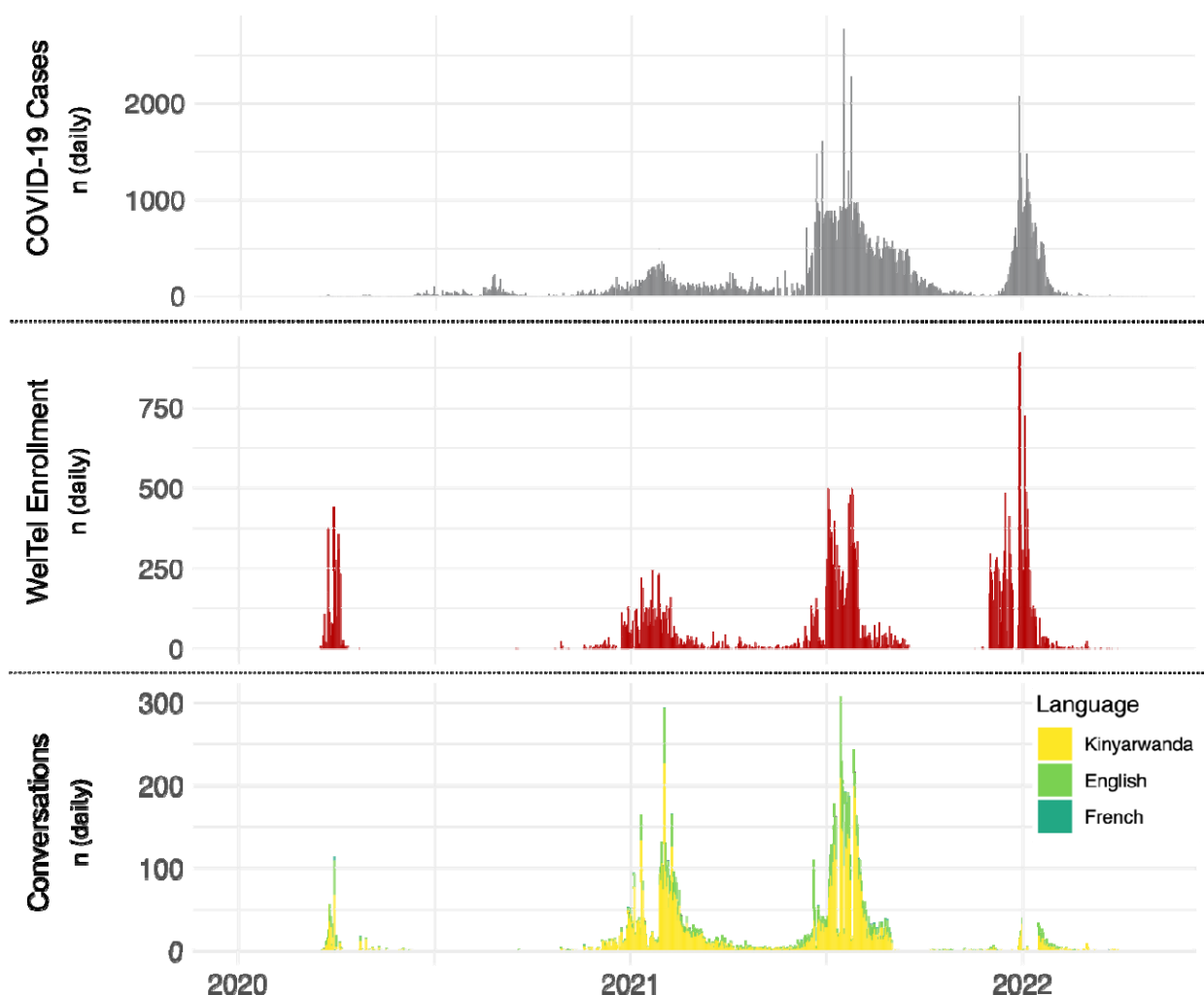


Figure 1. COVID-19 case incidence, and enrollment and use of the WelTel SMS text messaging service during the COVID-19 pandemic in Rwanda. *The grey plot captures the four major waves of case incidence in Rwanda. The red plot captures enrollment of COVID-19 patients (cases and contacts) in the WelTel system throughout the study period. Between June and September 2020, patients were not being enrolled in WelTel as the Rwandan government transitioned to government servers to host the software. The yellow and green plot captures the*

number of patient-provider SMS conversations throughout the study period; most conversations occurred in Kinyarwanda (67%) and English (25%) and other regional languages.

Of those registered, 6,021 patients (18%) used the texting service to generate 12,119 conversations (Table 1). Controlled for other demographic and clinical factors, the odds of generating a conversation among females compared to males was 0.83 (95%CI: 0.78-0.89; $p < 0.001$). Age did not significantly influence the frequency of conversations (OR per year increase in age = 1.00, 95%CI: 0.99-1.00, $p = 0.17$). Contacts were more likely to generate conversations than cases (OR 1.83 95%CI: 1.63-2.06, $p < 0.001$). The proportion registered patients texting conversations varied across waves of the pandemic (Wave one: 12%; two: 37%; three: 25%; and four: 3%).

Table 1. Characteristics of COVID-19 patients (cases and contacts) and usage of the WelTel text messaging service in Rwanda between March 18, 2020, and March 31, 2022. Percentages are within group. A proxy is an individual who allowed another/other individual(s) (e.g., a non-phone owner) to register for WelTel using their phone number (e.g., a mother could be a proxy for her non-phone-owning daughter). Shared refers to the individuals using a proxy (e.g., the daughter).

	People			Conversations		
	COVID-19 CASES	ENROLLED	PRODUCING CONVERSATIONS	CONVERSATION LENGTH		
				Median # Messages	IQR	Range
TOTAL	131,190	33,081	6,021	3	2-5	2-44
COVID-19 Status (at registration)						
Case	131,190 (100%)	30,398 (92%)	5,322 (88%)	3	2-5	2-44
Contact	-	2,683 (8%)	699 (12%)	4	3-8	2-30

Location of Residence						
Kigali City	42,152 (32%)	33,018 (97%)	5855 (97%)	3	2-5	2-44
Northern	20,347 (16%)	111 (0.3%)	22 (0.4%)	4	3-6	2-16
Southern	25,814 (20%)	370 (1%)	60 (1%)	3	2-5	2-18
Eastern	18,007 (14%)	233 (1%)	25 (0.4%)	5	3-7-5	2-11
Western	24,870 (19%)	161 (0.5%)	15 (0.2%)	3	2-4-5	2-16
Unknown	-	188 (0.6%)	44 (0.7%)	4	2-9	2-19
Sex						
Male	63,637 (49%)	17,102 (52%)	3,385 (56%)	3	2-5	2-36
Female	67,545 (51%)	14,665 (44%)	2,608 (43%)	3	3-5	2-44
Unknown	8 (0.0006%)	1,314 (4%)	28 (0.5%)	4	3-7	2-14
Age (mean (SD, range))	36.2 (17-36, 0-115)	33.9 (14-51, 0-100)	34.0 (13-27, 0-91)	-	-	-
Proxy/shared	-	1,726/2,749	453 (-)	-	-	-

The median number of conversations per patient was one (IQR: 1-2, range: 1-18), and the median conversation length (# of discrete text messages) was three (IQR: 2-5, range: 2-44). Most conversations occurred in Kinyarwanda (8,183, 67%) or English (3,069, 25%), however other regional languages (French, Kiswahili, Ganda, Chewa/Nyanja) were also identified (<10%). Language proportions were similar to the proportion of the population considered literate in those languages in Rwanda.²⁷

Comparison of NLP model performance

The NLP models tested classified most topics with a level of performance above our cut-off F1-score (Table 2). The traditional ML models performed better than fine-tuned transformer models on untranslated native language conversation data. Best performing models for each topic met the pre-determined F1-score cutoff of 0.7 for six of the nine

topics meeting the sample size threshold of 100. Model performance was roughly correlated with the number of occurrences (within sample) of the given topic.

Table 2. Performance of classification models on original language (untranslated), and English-translated conversations for each topic of interest. *Blue and orange fill indicate traditional-ML, and transformer architecture language models, respectively. Bold text indicates an F1 score meeting our performance cutoff (≥ 0.7). Topics are ordered by best-performing model F1 performance score (descending) when trained with original language (untranslated) conversations as ultimately used to label the full corpus.*

	Untranslated Conversations		English-translated Conversations	
Topic (n)	Classifier (Feature Extraction Method)	Performance (Precision Recall F1)	Classifier (Feature Extraction Method)	Performance (Precision Recall F1)
Diagnostic Methods (1030)	Random Forest (Character n-gram + TF-IDF)	0.93	BERT (NA)	0.95
		0.90		0.97
		0.92		0.96
Symptoms (2002)	KinyaBERT (NA)	0.87	BERT (NA)	0.92
		0.94		0.96
		0.91		0.94
Treatment (174)	KinyaBERT (NA)	0.85	Longformer (NA)	0.96
		0.88		0.96
		0.90		0.96
Prevention (459)	Ridge (Character n-gram + TF-IDF)	0.77	Longformer (NA)	0.97
		0.84		0.93
		0.81		0.95

Social (406)	Logistic Regression (Character n-gram + TF-IDF)	0.67	Longformer (NA)	0.88
		0.74		0.82
		0.70		0.85
Healthcare Logistics (281)	Logistic Regression (TF-IDF)	0.62	Ridge (Character n-gram + TF-IDF)	0.64
		0.81		0.88
		0.70		0.74
Service Quality (425)	Ridge (Character n-gram)	0.69	Longformer (NA)	0.83
		0.67		0.77
		0.68		0.80
Lifestyle/ Behavioural (156)	Logistic Regression (TF-IDF)	0.71	Longformer (NA)	0.92
		0.65		1.00
		0.68		0.96
Technical/ IT (185)	Logistic Regression (Non-Binary Bag of Words)	0.80	BERT (NA)	0.83
		0.43		0.54
		0.56		0.65

In our language experiments, fine-tuned transformer models generally performed better when using the English (translated) training and testing corpus, exceeding the F1 threshold in all but one topic.

Topic analysis of conversations

Human labeling (training corpus) and machine topic labeling (full corpus) identified the topics and subtopics of interest at roughly similar frequencies (Figure 2 and S3 Fig).

Overall, medical topics were discussed more frequently than non-medical topics. The most frequent topic was symptoms (69%), the majority reporting a lack of symptoms (59%). Diagnostic methods (38%, e.g., receiving COVID-19 test results), prevention (18%, including non-pharmaceutical such as how to maintain isolation, and pharmaceutical such as vaccines), healthcare logistics (16%, e.g., asking where to get tested), and treatment (8%, including management of symptoms). Social topics (19%) included mention of culture or religion (5.5%), followed by discussing friends and family (4.9%). Manually labeled training data revealed important issues such as nutrition and food security (3.7%). Examples of text conversations are provided in Table 3.

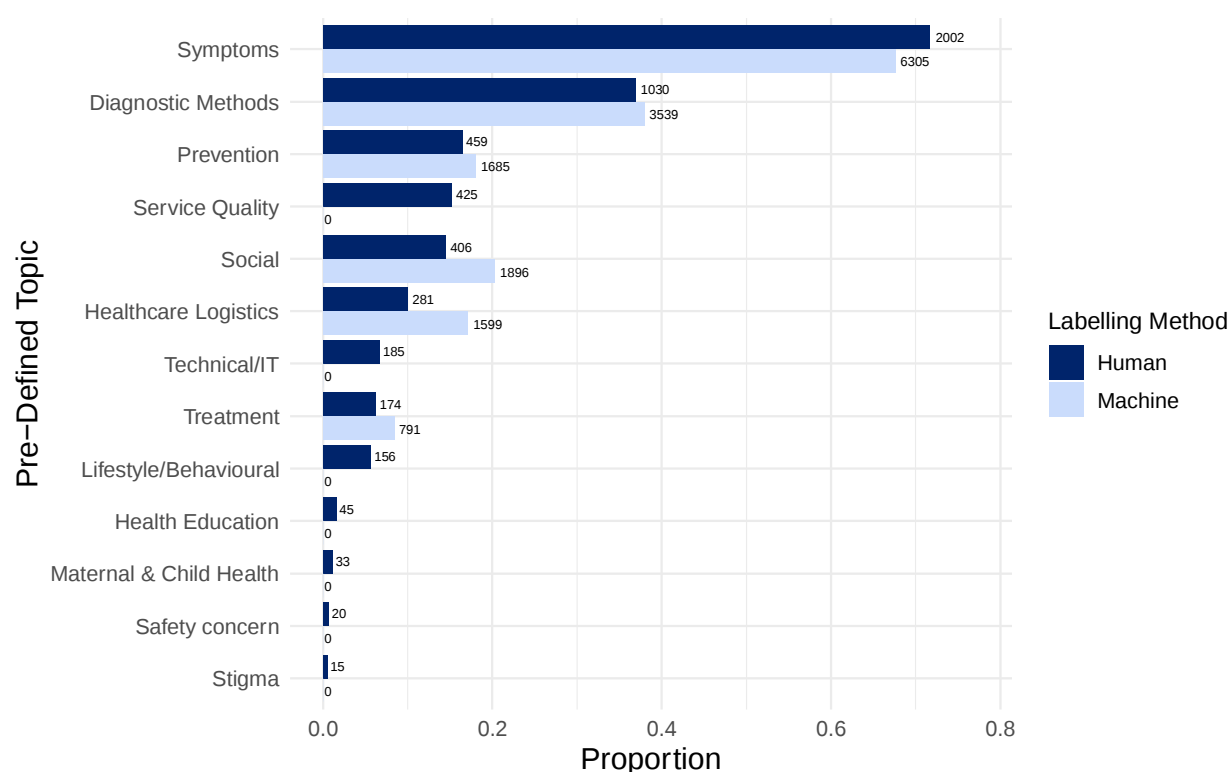


Figure 2. Proportion of conversations containing each topic of interest within the complete corpus (n=12,119) based on both human- (n=2,791) and machine-

labelling (n=9,328). *Note: conversations were often labelled with multiple topics/subtopics therefore proportions do not sum to 1. Some topics did not have sufficient data to produce models, or the models developed did not meet our performance cutoff of $F1 \geq 0.7$. For these, the proportion estimates are based upon human labelling alone.*

Comparing the odds of topic discussion within the full corpus by patient demographic and clinical factors revealed minor differences that are reported in (S4 Fig).

Table 3. Illustrative text conversations excerpts between public health officials and patients (cases and contacts) during the COVID-19 pandemic in Rwanda. “...” *indicates excluded text for brevity. Message types include: ‘(S):’ – a ‘system’ message containing customizable content with automated sending; ‘(P):’ – a ‘patient’ message containing patient remarks in free text form; and ‘(C):’ – a ‘clinician’ message containing clinician remarks in free text form.*

CONVERSATION TOPIC	EXAMPLE
Reporting a lack of symptoms	<i>(S): Hello, how are you today? Do you have any symptoms of coronavirus? [alongside Kinyarwanda and French translations of the message]</i> (P): We are ok. We do not have symptoms. <i>(C): Hello! We inform you that you were tested negative for coronavirus. Thank you.</i>
Reporting symptoms	<i>(S): Are you OK (Yes/No)? Or tell us if you have new symptom [alongside Kinyarwanda translation of the message]</i> (P): Good morning. I am coughing and I have flu but my situation is not alarming. I do not think that there is a problem.

	(C): Thank you for informing us. Explain your situation to the Doctor who is in charge of calling you. He is going to call you soon.
Diagnosis	(S): <i>haven't heard from you yet. how are you? [alongside Kinyarwanda translation of the message]</i> (P): Good morning. My name is #NAME#. I have been suffering from COVID-19 for days, but yesterday I went at #HEALTH FACILITY# to have a medical test for checking. Till now I have not yet got results via an sms. Can you help me to get those results if possible? (C): Good morning. We will inform people in charge to help you. Be patient. (C): #NAME# please note that you tested Negative on COVID-19 Ag RT on 2021-01-25. Code :*****
Treatment	(S): <i>Check-in message</i> (P): "...I took a Doliprane. Because I have little fever. 37.4..." (C): "Nice to hear that you are doing well. Doliprane helps for fever, body aches and headache..." (P): Yes, Thank you! Have a nice day
Pharmaceutical prevention	(S): <i>Check-in message</i> (P): "...Can you tell me where people suffering from heart issues and other diseases can go to get a vaccine?..." (C): "...Where do you usually go for a medical follow up...?" (P): "...#HEALTH FACILITY#" (P): I suffered from Covid 19 a month ago but now I am healthy. (C): You can contact them in order to know what they have scheduled for you. (P): Thank you
Non-pharmaceutical	(P): "...[Our] child has already tested positive...I have some

prevention	questions...Will not I be infected again once I share the same room and equipments with my child?...Is it possible to get a medical test for checking before ending 14 days of quarantine?..." (P): "...I am still suffering from coughing while the medicine I was taking is over..." (C): "...It is a must to go for a medical test for checking after 14 days. It would be better to clean your hands neatly, and to put on mask accordingly before taking your child..." (P): Thank you very much.
Social – food and nutrition	(P): "...I am suffering from COVID-19 and I live alone. I am starving and I don't have food. What can you help me? Thank you for your helpful answer." (C): Good afternoon. You can call the head of your village for a help. Food aid is being distributed these days. Thank you.
Social – culture and religion	<i>(S): Check-in message</i> (P): We are well this morning. We are really under the protection of Jesus. (C): That is good if you are well. Have a nice day.
Maternal & child health	<i>(S): Check-in message</i> (P): Hello. I am pregnant and I have tested positive for COVID-19. At the beginning I did not have any symptoms, but now I have many and I need your advice. (C): "...Be patient, you are going to get our support via a telephone call."
Gratitude	(C): Hello, do you have any health or COVID related concerns you would like to share with us today? (P): I tested negative for Covid 19. Thank you.

	(P): Thank you for your regular care. (P): Your advices are very important and we will always resort to them.
Safety concern	<i>(S): Check-in message</i> (P): We are in a very bad situation. ...We are mixed with people who have symptoms and those who do not have them, and we share the same bathroom and shower. Please, help us to get out of this situation before being infected. (C): Thank you for informing us. We are following up your situation. Keep on informing us. Thank you.
Proxy example	<i>(S): Check-in message</i> (P): Good morning. All of us, including children, are well. We don't have a problem. We hope to recover soon. Have a nice day. Thank you. (C): Thank you and have a nice day too.

Methods

Setting and digital health monitoring technology

Rwanda, a Central African nation with a population of 14 million, was internationally recognized for its prompt response to the COVID-19 pandemic.¹⁷ Their public health command implemented the WelTel text messaging service [www.welitelhealth.com] on March 18, 2020, coincident with detecting its first COVID-19 cases. This tool enabled

public health staff to remotely monitor and support individuals who tested positive for, or were at high risk of contracting, SARS-CoV-2, and remained operational until the public health emergency was officially declared over.

The WelTel platform is a secure web-based application that healthcare providers can access from any internet-connected device. It interfaces with cellular networks enabling it to send both pre-set and manually input SMS messages to registered mobile phone users. It features a conversation dashboard that visualizes responses.

Following the initial facility-based quarantine of COVID-19 cases identified at border entry points, Rwanda launched its HBC program when community transmission became widespread. Under this program, mild cases and contacts were asked to stay home and be remotely monitored. Patients with significantly worsening clinical status were triaged to hospitals. Individuals were enrolled as 'cases' if they tested positive for SARS-CoV-2 or as 'contacts' if they were deemed high risk due to exposure. Patients without personal cell phone numbers were registered via a household member, friend, or neighbor.

The messaging protocol for COVID-19 monitoring involved sending bulk automated daily 'check-in' texts in Rwanda's principal languages (Kinyarwanda, English, French) daily to registered patients throughout their isolation period. Patients could respond via text in their own words to indicate their status and/or ask questions. Replies indicating a problem or question were flagged for follow-up. Public health clinicians responded to

flagged messages to provide information or advice, forming interactive text conversations.

Included datasets and description

Patient registration and all text communication data generated between March 18, 2020 and March 31, 2022 was exported from the WelTel database and linked to the Rwanda DHIS2 COVID-19 testing registry to obtain missing demographic or clinical data.¹⁸ Personal identifiers were removed. For the purposes of our study, a ‘conversation’ was defined as a text exchange consisting of at least one ‘incoming’ patient message containing three or more words (filtering out minimal replies e.g., ‘doing well’), together with at least one ‘outgoing’ public health clinician or automated system message (thus being interactive). Patient registration and numbers of conversations and their characteristics (# messages, language used) were quantified. Conversation usage was compared across patient sociodemographic and clinical characteristics (sex, age, province of residence, pandemic wave, and COVID-19 status (case vs. contact)) using a multivariable logistic regression model. Odds ratios (OR) are reported with 95% confidence intervals (CI) and $p < 0.05$ was considered significant. Example conversations were selected to illustrate context.

Language translation and topic labelling

A list of topics and subtopics was developed through a three-round mini-Delphi process involving ten public health experts, clinicians, and researchers from Rwanda and Canada. This list was an expansion of a previously published list focused on clinical care and used to develop a conversation analysis and visualization tool, ConVIScope.¹² The final list of topics and subtopics encompassed perspectives on public health, emergency response, and clinical care. Topic definitions are provided in (S1 Fig).

For the purpose of labeling a training dataset for topic classification models, a sample of 2,791 text conversations from March 18, 2020, to May 31, 2021, was extracted from the study period corpus. These conversations were analyzed for language use with Google's CLD2 language detector. Non-English sample messages were translated into English by a compensated professional healthcare text translator proficient in Kinyarwanda, English, French, and Kiswahili.

The English-translated conversations were labeled by at least three trained members of the research team using a custom conversation annotation tool. Any conflicts in labeling were resolved by an external team member. The topic labels were then applied back to the corresponding conversations in their original language, where applicable.

Classification model training, testing, and application

We aimed to identify the best performing classification model for each topic and subtopic. Training set sample size constraints limited our classification experiments to

topics with ≥ 100 occurrences. We employed the binary relevance strategy, which simplifies performance management by building a separate binary classifier for each topic. We experimented with both traditional ML algorithms and Transformer-based language models for each topic. We explored large language models (LLMs, e.g., GPT-4), however, our literature review revealed that, of the three pre-trained with any Kinyarwanda, the pre-training datasets for these models contained minimal Kinyarwanda text ($\leq 0.01\%$). Investigations of the performance of these LLMs at NLP tasks in Kinyarwanda suggested they were insufficient to include.¹⁹

Traditional-ML classification models: For training traditional ML models, we explored five feature sets: binary and non-binary bag of words, TF-IDF (term-frequency inverse document frequency), and character n-gram (with/without TF-IDF). The character n-gram features were tried because Kinyarwanda is an agglutinative language. All combinations of one of these feature sets, and one classification algorithm, were tested. We evaluated logistic regression, ridge classifier, and random forest as classification algorithms. Hyperparameters for each classifier were tuned using random search in five-fold cross validation.²⁰

Transformer-Based Language Models: We leveraged open-source pre-trained Transformer-based language models for Kinyarwanda, including AfriBERTa, AfroLM, AfroXLMR, and KinyaBERT.^{21–24} To fine-tune these models for our task, we employed the AdamW optimizer along with the binary cross-entropy loss function. Additionally, we incorporated dropout layers (with a probability of 0.0 for the first layer and 0.3 for subsequent layers) for regularization.

We compared the performance of each model for each topic using the validation set F1-score, with a minimum F1-score cut-off of 0.7. The highest F1-score models were applied to predict topics for the remaining 9,328 unlabeled, untranslated SMS conversations.

Additional Language experiments: Since our corpus contained a mix of high- and low-resource languages (e.g., English, French, Kinyarwanda) we explored how this might affect performance of the models by conducting identical experiments to above using only the English-translated version of the training corpus. We evaluated both traditional ML models (as above), and English pre-trained Transformer models (BERT, Longformer).^{25,26} F1 scores are reported for comparison.

See (S2 Fig.) for more hyperparameter and fine-tuning details.

Topic analysis and statistics

Counts and proportions of conversations containing each topic (full corpus), and subtopic (sampled corpus), are provided. Topic frequencies were compared by patient demographics and clinical metrics using multiple independent multivariable logistic regression models for each topic. Models incorporate all patient demographic and clinical characteristics and use complete cases. OR are reported with 95% CIs and p-values are adjusted for multiple testing using the Benjamini-Hochberg correction with a false discovery rate of 0.05. Classified topic categories could be batch filtered and

manually read to understand the nature of the discussions that occurred within each topic.¹² Data summarizations, visualizations, and statistical testing were produced or conducted using R version 4.2.1.

Discussion

To our knowledge, this is the first study to report on the use of NLP to investigate a broad range of topics from a complete corpus of conversational texts with patients in home isolation in a pandemic or outbreak setting at large scale. Through a combination of manual annotation and NLP techniques, we gained insights into patient experiences during COVID-19 in Rwanda by identifying key discussion topics between patients and public health clinicians. The sustained use of the text messaging service in Rwanda demonstrated the feasibility of using SMS-based communication for remote monitoring at scale in a region with limited internet access but widespread mobile phone penetration. Leveraging existing infrastructure allowed for efficient support for isolated individuals, regardless of location or access to healthcare facilities.

Our data captured the first two years of the Rwandan COVID-19 pandemic, including text conversations in Kinyarwanda, English, and other local languages with public health clinicians to report their status or seek advice. Topic classification revealed that medical topics, such as symptoms, diagnostics, prevention, and treatment, were commonly discussed, reflecting patient's primary focus on their health status and obtaining medical guidance. However, social and lifestyle topics, including discussions about social

support, healthcare logistics, diet/nutrition, finances, and service quality, also emerged as important aspects to patients during home isolation. The data underscores the need for holistic support systems for patient isolation programs. Additionally, while our topic classification approach rapidly quantified prevalent topics well, reflecting their relative importance to patients, it also enables the sorting of large datasets into smaller categories for easier manual evaluation by decision-makers. The identification of less common topics, such as food insecurity, may also provide actionable targets for public health responders, especially if it can be localized to communities.

NLP techniques are rapidly being introduced into health systems to understand unstructured data from large datasets, such as clinical notes in electronic medical records (EMRs), social media posts, or free-text survey responses.¹¹ For example, topic classification has been employed to sort unstructured patient feedback from text messages.¹⁵ Still, manual qualitative analyses dominate. A manually labeled thematic analysis of 1,454 patient portal secure messaging conversations between cancer patients and health professionals in the US during COVID-19 found 26% of conversations were related to COVID-19 and identified topics related to changes in care plans, symptoms, risks, precautions and symptoms.²⁸ In Northern Thailand manual topic classification of SMS exchanges with 213 home-isolated patients with diagnosed COVID-19 revealed several complexities of care including discussing symptoms or vital signs, medications, diet and exercise, mental stresses, information about COVID-19, and a high proportion of logistical assistance to navigate the health services.²⁹ Manual labeling is resource intensive and generally limited to smaller data sets. One US-based

study used NLP techniques to determine if patient-initiated secure messaging via a EMR portal was associated with positive testing, but this was a simple binary classification.¹³ To our knowledge, our study is the first to use NLP techniques to classify patient clinician texts into broad medical, social and logistical topics set by experts, and at a national scale.

Our study compared performance of several ML model approaches. Traditional ML models and transformer-based language models were both effective in classifying text message conversations into the predefined topics. While traditional ML models performed better with untranslated conversations, transformer models had superior performance when trained on English (translated) data, highlighting their limited performance on African languages.³⁰ This observation is consistent with the knowledge that transformer architecture language models are likely to underperform on NLP tasks in low-resource languages until sufficient high-quality pre-training data is sourced and leveraged.³¹ Similarly, LLMs, which are also based on transformer architecture, currently contain too little, if any, training on Kinyarwanda and other low-resource languages to be useful in our context.¹⁹ A brief trial using Google Translate to machine translate Kinyarwanda text into English for our study was inaccurate. This highlights the importance of considering language-specific model performance when analyzing multilingual text data.³² Language consideration remains a significant challenge, with over 4,000 written languages worldwide. Pandemics affect everyone and AI tools must account low-resource languages to be equitable.

Our study has limitations. First, the supervised learning approach required the pre-definition and manual labeling of topics of interest which can be human resource intensive. Furthermore, the topics of interest or training samples annotated may evolve over time or change in different settings requiring continuously updating training datasets. For example, prevalent discussions of vaccines came later in the pandemic after most of our training data had been annotated. Including unsupervised ML models that discover topics as they emerge could help overcome this bias, but have their own flaws. Furthermore, while our dataset was similar in size to those used to train other topic classification models in healthcare, many subtopics of interest were not prevalent enough to develop models. Additionally, the training dataset was translated into English for labelling purposes, again, being resource intensive and potentially altering meaning in translation. Finally, training the models took time and conducting the NLP analyses afterwards meant the data was not available to decision-makers real time. Ideally, once models are trained and validated, they could be applied while an outbreak or pandemic is occurring and provide insights into patient and population issues that could be responded to while they are occurring. Together with location data, this type of analysis could contribute to precision public health responses.

Conclusion

Moving forward, future research should explore ways to further optimize the evaluation of text-based communication for greater insights into the patient care journey in pandemics, and clinical care. It must consider technology infrastructure available to

populations globally and be inclusive of resource-limited languages.^{33,34} Efforts to improve language support and translation capabilities across diverse linguistic and cultural contexts will be essential to ensure equitable access to digital services by people in all regions. Ultimately, emerging research on models that detect, understand, and generate natural language, as well as development of tools that formulate predictions and recommendations for healthcare decision-makers, will achieve the higher accuracy, precision, and reliability required for use in routine healthcare.

Acknowledgements

We would like to thank the staff at the Rwanda Biomedical Center and work-learn students at the University of British Columbia who helped with language translation, annotation, and various aspects of background research and model building.

References

1. Reddy KP, Shebl FM, Foote JHA, Harling G, Scott JA, Panella C, et al. Cost-effectiveness of public health strategies for COVID-19 epidemic control in South Africa: a microsimulation modelling study. *Lancet Glob Health*. 2021 Feb;9(2):e120–9.
2. Mbunge E, Fashoto S, Akinuwesi B, Gurajena C, Metfula A. Challenges of Social Distancing and Self-Isolation during COVID-19 Pandemic in Africa: A Critical Review [Internet]. Rochester, NY; 2020 [cited 2024 Jun 13]. Available from: <https://papers.ssrn.com/abstract=3740202>
3. Wang F, Gao Y, Han Z, Yu Y, Long Z, Jiang X, et al. A systematic review and meta-analysis of 90 cohort studies of social isolation, loneliness and mortality. *Nat Hum Behav*. 2023 Aug;7(8):1307–19.
4. Babili A, Nsanzimana S, Rwagasore E, Lester RT. SMS-based digital health intervention in Rwanda's home-based care program for remote management of

COVID-19 cases and contacts: A qualitative study of sustainability and scalability.
Front Digit Health. 2023 Jan 9;4:1071790.

5. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a
mobile phone short message service on antiretroviral treatment adherence in
Kenya (WelTel Kenya1): a randomised trial. The Lancet. 2010 Nov
27;376(9755):1838–45.

6. Rathgeber SL, Hutchison SM, De Souza AM, Lester R, Blydt-Hansen T, Human
DG, et al. A text messaging intervention and quality of life in adolescents with solid
organ transplants. Pediatr Transplant. 2022;26(3):e14219.

7. van der Kop ML, Muhula S, Nagide PI, Thabane L, Gelmon L, Awiti PO, et al.
Effect of an interactive text-messaging service on patient retention during the first
year of HIV care in Kenya (WelTel Retain): an open-label, randomised parallel-
group study. Lancet Public Health. 2018 Mar;3(3):e143–52.

8. MAIN INDICATORS: 5th Rwanda Population and Housing Census (PHC) |
National Institute of Statistics Rwanda [Internet]. [cited 2024 Jun 13]. Available
from: https://statistics.gov.rw/publication/main_indicators_2022

9. Nachega JB, Leisegang R, Kallay O, Mills EJ, Zumla A, Lester RT. Mobile Health
Technology for Enhancing the COVID-19 Response in Africa: A Potential Game
Changer? Am J Trop Med Hyg. 2020 Jul;103(1):3–5.

10. Kop ML van der, Karanja S, Thabane L, Marra C, Chung MH, Gelmon L, et al. In-Depth Analysis of Patient-Clinician Cell Phone Communication during the WeTel Kenya1 Antiretroviral Adherence Trial. PLOS ONE. 2012 Sep 25;7(9):e46033.
11. Zhou B, Yang G, Shi Z, Ma S. Natural Language Processing for Smart Healthcare. IEEE Rev Biomed Eng. 2024;17:4–18.
12. Li R, Hoque E, Carenini G, Lester R, Chau R. ConVIScope: Visual Analytics for Exploring Patient Conversations [Internet]. arXiv; 2021 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/2108.13514>
13. Mermin-Bunnell K, Zhu Y, Hornback A, Damhorst G, Walker T, Robichaux C, et al. Use of Natural Language Processing of Patient-Initiated Electronic Health Record Messages to Identify Patients With COVID-19 Infection. JAMA Netw Open. 2023 Jul 7;6(7):e2322299.
14. Lu H, Ehwerhemuepha L, Rakovski C. A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. BMC Med Res Methodol. 2022 Jul 2;22(1):181.
15. Cronin RM, Fabbri D, Denny JC, Rosenbloom ST, Jackson GP. A comparison of rule-based and machine learning approaches for classifying patient portal messages. Int J Med Inf. 2017 Sep;105:110–20.
16. Liu X, Alderman J, Laws E. A Global Health Data Divide. NEJM AI. 2024 May 23;1(6):Ale2400388.

17. Condo J, Uwizihiwe JP, Nsanzimana S. Learn from Rwanda's success in tackling COVID-19. *Nature*. 2020 May 26;581(7809):384–384.
18. Krafft M. Going paperless for COVID-19 testing in Rwanda with DHIS2 Android Capture App [Internet]. DHIS2. 2020 [cited 2024 Jun 13]. Available from: <https://dhis2.org/rwanda-covid-testing/>
19. Moslem Y, Haque R, Kelleher JD, Way A. Adaptive Machine Translation with Large Language Models [Internet]. arXiv; 2023 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/2301.13294>
20. Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. *J Mach Learn Res*. 2012;13(10):281–305.
21. Ogueji K, Zhu Y, Lin J. Small Data? No Problem! Exploring the Viability of Pretrained Multilingual Language Models for Low-resourced Languages. In: Ataman D, Birch A, Conneau A, Firat O, Ruder S, Sahin GG, editors. *Proceedings of the 1st Workshop on Multilingual Representation Learning* [Internet]. Punta Cana, Dominican Republic: Association for Computational Linguistics; 2021 [cited 2024 Jun 13]. p. 116–26. Available from: <https://aclanthology.org/2021.mrl-1.11>
22. Dossou BFP, Tonja AL, Yousuf O, Osei S, Oppong A, Shode I, et al. AfroLM: A Self-Active Learning-based Multilingual Pretrained Language Model for 23 African Languages [Internet]. arXiv; 2022 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/2211.03263>

23. Alabi JO, Adelani DI, Mosbach M, Klakow D. Adapting Pre-trained Language Models to African Languages via Multilingual Adaptive Fine-Tuning [Internet]. arXiv; 2022 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/2204.06487>
24. jean-paul/KinyaBERT-small · Hugging Face [Internet]. [cited 2024 Jun 13]. Available from: <https://huggingface.co/jean-paul/KinyaBERT-small>
25. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Internet]. arXiv; 2019 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/1810.04805>
26. Beltagy I, Peters ME, Cohan A. Longformer: The Long-Document Transformer [Internet]. arXiv; 2020 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/2004.05150>
27. Rwanda's literacy rate rises | National Institute of Statistics Rwanda [Internet]. [cited 2024 Jun 13]. Available from: <https://statistics.gov.rw/publications/article/rwanda%E2%80%99s-literacy-rate-rises>
28. Alpert JM, Campbell-Salome G, Gao C, Markham MJ, Murphy M, Harle CA, et al. Secure Messaging and COVID-19: A Content Analysis of Patient-Clinician Communication During the Pandemic. Telemed J E-Health Off J Am Telemed Assoc. 2022 Jul;28(7):1028–34.
29. Pinyopornpanish K, Nantsupawat N, Buawangpong N, Pliannuom S, Vaniyapong T, Jiraporncharoen W. Concerns of Home Isolating COVID-19 Patients While Receiving Care via Telemedicine during the Pandemic in the Northern Thailand: A

Qualitative Study on Text Messaging. Int J Environ Res Public Health. 2022 May 28;19(11):6591.

30. Wu J, Liu X, Li M, Li W, Su Z, Lin S, et al. Clinical Text Datasets for Medical Artificial Intelligence and Large Language Models — A Systematic Review. NEJM AI. 2024 May 23;1(6):Alra2400012.

31. Hedderich MA, Lange L, Adel H, Strötgen J, Klakow D. A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios [Internet]. arXiv; 2021 [cited 2024 Jun 13]. Available from: <http://arxiv.org/abs/2010.12309>

32. Yang LWY, Ng WY, Lei X, Tan SCY, Wang Z, Yan M, et al. Development and testing of a multi-lingual Natural Language Processing-based deep learning system in 10 languages for COVID-19 pandemic crisis: A multi-center study. Front Public Health. 2023 Feb 13;11:1063466.

33. Adebamowo C, Callier S, Maduka OC, Akintola S, Kukucka J, Arima CG, et al. Ethical Oversight of Data Science Health Research in Africa. NEJM AI. 2024 May 23;1(6):Alpc2400033.

34. Serra-Burriel M, Locher L, Vokinger KN. Development Pipeline and Geographic Representation of Trials for Artificial Intelligence/Machine Learning–Enabled Medical Devices (2010 to 2023). NEJM AI. 2023 Dec 11;1(1):Alp2300038.

Supporting Information

S1 Figure. Defined *topics and subtopics of interest were determined by a Delphi process*
contributed by expert members of the study team and incorporated into a bespoke annotation
software used to label conversations.

Topic	Subtopic	Definition and Examples
Symptoms	Physical	Any physical symptom, sensation, or feature (pleasant and unpleasant) that is mentioned in the conversation. Examples: pain, weight loss, burning, fatigue, shortness of breath, fever, diarrhea, muscle aches, cramping, coughing, bleeding, broken arm etc.
	Mental / Emotional	Any mental/emotional feeling, sensation, or feature (pleasant and unpleasant) that is mentioned in the conversation. Examples: confusion, depression, fear, anxiety, low energy, extreme mood, suicidal ideation, feelings of guilt, lack of interest, stigma etc.
	No Symptoms	Patient explicitly mentions not having symptoms. Examples: I have no symptoms, I feel fine.
Diagnostic Methods	Clinical	The process of identifying a disease, condition, or injury based on the signs and symptoms a patient is having and the patient's health history and physical exam.
	Laboratory/Testing	Discussion related to tests on blood, urine, or other tissues or substances taken from the body to help diagnose disease or other conditions, whether performed or not. Examples: blood tests (CBC, FBG, Hb1Ac, TSH, lipid panel, etc.), blood cultures, urinalysis, pulmonary function tests, biopsies, stool test (FIT, FOBT), etc.
	Imaging	Studies that provide a picture of the body's interior to help diagnose

		disease or other conditions. Examples: x-ray, CT, MRI, fMRI, ultrasound, PET scan, etc.
	Other	Other diagnostic methods/topic(s) not listed.
Treatment (Rx)	Medications	Pharmaceutical drug that usually requires a medical prescription to be dispensed. Examples: antibiotics, antidepressants, anti-hypertensives, amoxicillin, escitalopram, ramipril, etc. include over-the-counter medications (e.g., Tylenol)
	Procedures	Surgery, Breathing procedures, Radiation Therapy etc.
	Alternative	Non-traditional method of healing/treating disease/condition. Examples: herbal medicine, acupuncture, massage therapy, naturopathy, meditation, etc.
	Physical Therapy	Care that aims to ease pain and helps function, move, and live better. Examples: physiotherapy (geriatric, neurological, orthopedic, chiropractic etc.)
	Counselling	Giving advice/coaching/making recommendations to the patient (can be non-health related). Examples: psychologic counseling, coaching of lifestyle modifications like diet beyond standard clinical advice (e.g., in-depth discussion on how to reduce salt intake, not simply saying to reduce salt), smoking cessation, the importance of exercise on cardiovascular health, etc.
	Other	Other treatment topic(s) not listed.
Prevention	Pharmaceutical	Any mention of an intervention that utilizes chemical or biological

	Prevention	methods given to or taken by humans in an attempt to prevent disease. Examples: prophylactic medications, immunizations, treatment as prevention, etc.
	Non-Pharmaceutical Prevention	Any mention of an intervention that does NOT rely on bio/chemical products given to or taken by humans in an attempt to prevent disease. Especially those methods that are behavioural (ex: physical distancing) or structural (ex: laws or policies). Examples: personal protective equipment (e.g., masking), physical distancing, barriers, quarantine/isolation, lockdowns, school closures, education programs, policies, laws, testing without treatment, environmental decontamination, animal/pest control, etc.
Healthcare Logistics	Outpatient logistics/scheduling	Mention or pertaining to setting up or attending outpatient clinical care or diagnostics. Examples: clinic, family doctor office, ECG appointment, hemodialysis appointment, dentist appointment, specialist appointment, physiotherapy, etc.
	Hospitalization	Mention or pertaining to any admission or stay in hospital. Receiving medical treatment during hospital admission. Examples: visit to emergency department, or regarding planning, admission, or discharge from hospital care. Includes long-term care facilities.
Lifestyle/Behavioural	Diet / Nutrition	Relating to food/liquid intake and health. Examples: nutrition related to weight management, apple, fruits, pie, sandwich, water, Coca-Cola, dinner, lunch, waffles, chips, popcorn,

		fish, steak, etc.
	Exercise	Relating to physical activity done to improve health and fitness. Examples: working out, going to the gym, jogging, playing soccer, playing basketball, doing yoga, walking, walking the dog, etc.
	Substance Use	Relating to the use of drugs or alcohol, and includes substances such as cigarettes, illegal drugs, prescription drugs, inhalants, solvents, etc. Examples: alcohol, marijuana, tobacco use, fentanyl, ecstasy, cocaine, addiction of substances etc.
	Other	Other lifestyle topic(s) not listed.
Social/Environmental	Housing	Relating to housing, shelter. Examples: "living at home currently," "bought a new house," "moving into an apartment in Victoria," "can't afford a place," "got evicted," "am homeless".
	Work/School	Relating to work or school. Examples: "studying psychology," "UBC," "lots of studying to do," "stressed from exams," "I'm a banker," "work was so busy," "got fired," etc.
	Social Services	Relating to government services provided for the benefit of the community, such as education, medical care, and housing programs. E.g.: family benefits, welfare, income assistance, housing assistance, food subsidies, foster care, disability support, mental

		health services, etc.
	Friends & Family	Relating to friends and family relations. E.g.: "went to see my mom," "spent time with friends," " hiking with my partner," "family dinner," "hung out with my best friend,"
	Cultural/Religion	Relating to ideas and customs of a culture. Examples: "Chinese herbal medicine," "fasting for Ramadan," "sweat lodges," religious and spiritual practices, cultural institutions i.e., church, having own "ideas" about their own illness based on their culture, refusing conventional treatment because of what is taught in their culture.
	Travel	Mention or related to transportation and or travel. E.g., vacation in Mexico, work abroad, traveled to... I'm currently in Toronto.
	Physical Environment/Climate	Mention or pertaining to the climate or physical environment. E.g., smoke from fires, sun exposure, cold related illness, concerns re logging or farming practices
	Financial	Mention or pertaining to personal or other finances. E.g., cost of something, affordability, economics.
	Other	Other social topic(s) not listed.
Service Quality	Problems Solved	An attempt that addresses the patient's problem within the context of one conversation, regardless of if it was ultimately effective, but at least addressed/attempted to address a present issue
	Grateful Patient	Demonstration of gratitude toward the HCP or care provided.

		Examples: Thanks! (if appears as genuine gratitude, beyond routine customary formalities). I really appreciate. (NOT: See you later, thanks!)
	Service Complaint	Conversation contains any concern or complaint brought forward about the care one is receiving, even if it also contains positive mentions.
	Request to Stop	Select this box if, at any time, the patient requests to be removed from the messaging platform or stop receiving messages.
Safety Concerns	No Subtopics	Discussing worry or concern about their life or the life of others based on specific circumstances. Example: I am worried about quarantining in a healthcentre where people are contagious.
Technical/IT	No Subtopics	Non-healthcare technical issues. Examples: trouble with texting or health IT, video conferencing and virtual care access
Maternal & Child Health	No Subtopics	Any mention or pertaining to pregnancy, breastfeeding, motherhood, and early childhood care
Health Education	No Subtopics	Providing information to help patients beyond simple advice, with the intent to increase knowledge - can be verbal, written or in the form of visual aids like pamphlets, courses. Examples: pamphlet on what diabetes is, teaching how to self-monitor blood pressure, etc.
Stigma	No Subtopics	A mention of the disapproval of, or <u>discrimination</u> against, an individual or group based on perceivable characteristics that serve to distinguish them from other members of a society. Social stigmas are commonly related to <u>culture</u> , <u>gender</u> , <u>race</u> , <u>socioeconomic class</u> , age, sexual orientation, <u>body image</u> , intelligence or lack

thereof, and health. Can be self-directed.

S2 Figure. NLP Model Details

Traditional machine learning feature extraction details						
Feature Extractor		Detail				
Character n-gram		Experimented with word bound character ranges of 1-3, 1-4, and 1-5.				
Traditional machine learning model hyperparameters and random search ranges						
Model		Hyperparameter		Range		
Logistic Regression		C (inverse of regularization strength)		Loguniform (1e-4, 1e4)		
		Penalty		L1, L2		
Random Forest		Max_features (number of features to consider when looking for best split)		Uniform (0,1)		
		Min_samples_leaf (minimum number of samples required to be at a leaf node)		Randint (1, 20)		
Ridge Classifier		Alpha (regularization strength)		Loguniform (1e-2, 1e3)		
Transformer model hyperparameters						
Model	Epochs	Learning Rate		Loss	Optimizer	Batch Size
AfriBERTa ^a	15	Starts at 0.01*		BCEWithLogitsLoss	AdamW	32
AfroXLMR ^b	15					
AfroLM ^c	15					
KinyaBERT ^d	15	10 ⁻⁵				

* We utilized the ReduceLROnPlateau scheduler that reduces the learning rate by a factor of 0.5 for every two epochs where the validation loss does not decrease, up to a minimum value of 10⁻⁵.

^a <https://aclanthology.org/2021.mrl-1.11/>

^b <https://arxiv.org/abs/2204.06487>

^c <https://arxiv.org/abs/2211.03263>

^d <https://huggingface.co/jean-paul/KinyaBERT-small>

S3 Figure. Frequency and proportion of conversations in which all topics and subtopics of interest occurred within the human-labelled, machine-labelled, and complete corpora. Note: Conversations were often labelled with multiple topics/subtopics, therefore percentages do not sum to 100%.

	Human Labelled (n=2,791)	Machine Labelled (n=9,328)	Human-and-Machine-Labelled (n=12,119)
Topic Subtopic(s)	n (%)	n (%)	n (%)
Symptoms	2,002 (72)	6,305 (68)	8,307 (69)
None	1,657 (59)	-	-
Physical	406 (15)	-	-
Mental/Emotional	45 (1.6)	-	-
Diagnostic Methods	1,030 (37)	3,539 (38)	4,569 (38)
Laboratory/Testing	1,018 (36)	-	-
Imaging	7 (0.2)	-	-
Clinical	5 (0.2)	-	-
Other	5 (0.2)	-	-
Prevention	459 (16)	1,685 (18)	2,114 (18)
Pharmaceutical	17 (0.6)	-	-
Non-pharmaceutical	443 (16)	-	-

Service Quality	425 (15)	NP	NP
Grateful Patient	244 (8·7)	-	-
Service Complaint	106 (3·8)	-	-
Request to Stop	50 (1·8)	-	-
Problems Solved	45 (1·6)	-	-
Social	406 (15)	1,896 (20)	2,302 (19)
Cultural/Religion	154 (5·5)	-	-
Friends & Family	138 (4·9)	-	-
Financial	53 (1·9)	-	-
Work/School	46 (1·7)	-	-
Housing	32 (1·2)	-	-
Social Services	30 (1·1)	-	-
Travel	12 (0·4)	-	-
Environment/Climate	6 (0·2)	-	-
Other	7 (0·3)	-	-
Healthcare Logistics	281 (10)	1,599 (17)	1,880 (16)
Logistics/Scheduling	268 (9·6)	-	-
Hospitalization	15 (0·5)	-	-
Technical/IT	185 (6·6)	NP	NP
Treatment	174 (6·2)	791 (8·5)	965 (8·0)
Medications	152 (5·5)	-	-
Alternative	35 (1·2)	-	-

Physical Therapy	1 (0.04)	-	-
Procedures	0 (0)	-	-
Counseling	0 (0)	-	-
Other	0 (0)	-	-
Lifestyle/Behavioural	156 (5.5)	NP	NP
Diet/Nutrition	102 (3.7)	-	-
Exercise	25 (0.9)	-	-
Substance Use	0 (0)	-	-
Other	52 (1.9)	-	-
Health Education	45 (1.6)	NT	NT
Maternal & Child Health	33 (1.2)	NT	NT
Safety Concern	20 (0.7)	NT	NT
Stigma	15 (0.5)	NT	NT

NP = Not Predictable (best model did not meet F1 performance cutoff and was not applied)

NT = Not Tested (insufficient data to develop models)

S4 Table. Comparison of the odds of discussion of predictable conversation topics of interest (i.e., those conversation topics of interest for which we developed a classification model of interest meeting the $F1 \geq 0.7$ performance cutoff) by patient demographic and clinical factors in the complete conversation corpus ($n=12,119$). Presented p-values were adjusted for multiple testing using the Benjamini-Hochberg correction.

TOPIC	COMPARISON	ODDS RATIO (95% CI, adj. p-value)
-------	------------	-----------------------------------

	(vs. reference)	
Symptoms	Female vs. Male	1.04 (0.96-1.13, p=0.39)
	Age (per year increase)	1.00 (1.00-1.00, p=0.47)
	Contact vs. Case	1.25 (1.06-1.47, p=0.02)
Diagnostic Methods	Female vs. Male	0.87 (0.80-0.94, p=0.0012)
	Age (per year increase)	1.00 (0.99-1.00, p=0.060)
	Contact vs. Case	0.84 (0.73-1.00, p=0.060)
Social	Female vs. Male	1.04 (0.94-1.14, p=0.45)
	Age (per year increase)	1.005 (1.002-1.009, p=0.012)
	Contact vs. Case	1.09 (0.91-1.30, p=0.43)
Prevention	Female vs. Male	0.93 (0.84-1.02, p=0.29)
	Age (per year increase)	1.00 (0.99-1.00, p=0.22)
	Contact vs. Case	0.61 (0.49-0.75, p=0.00004)
Healthcare Logistics	Female vs. Male	0.88 (0.79-0.98, p=0.025)
	Age (per year increase)	1.005 (1.0009-1.008, p=0.025)
	Contact vs. Case	0.69 (0.55-0.86, p=0.003)
Treatment	Female vs. Male	1.077 (0.94,1.23, p=0.36)
	Age (per year increase)	1.006 (1.0009-1.011, p=0.048)
	Contact vs. Case	0.30 (0.20-0.46, p<0.0001)

579

580

