RESEARCH ARTICLE

Using Machine Learning to Establish the Concerns of Persons With HIV/AIDS During the COVID-19 Pandemic From Their Tweets

RICHARD K. LOMOTEY [1], (Member, IEEE),
SANDRA KUMI [2], (Graduate Student Member, IEEE), **MAXWELL HILTON** [1],
RITA ORJI [3], (Member, IEEE), AND **RALPH DETERS** [2], (Member, IEEE)

[1]Department of Information Sciences and Technology, The Pennsylvania State University—Beaver, Monaca, PA 15061, USA
[2]Department of Computer Science, University of Saskatchewan, Saskatoon, SK S7N 5C9, Canada
[3]Faculty of Computer Science, Dalhousie University, Halifax, NS B3H 4R2, Canada

Corresponding author: Richard K. Lomotey (rkl5137@psu.edu)

ABSTRACT There are millions of People Living with HIV/AIDS (PLWHA) globally and over the years, addressing their concerns has been topical for many stakeholders. It is a well-known and established fact that PLWHA are at increased risk of victimization and stigmatization. Unfortunately, the world experienced an outbreak of the COVID-19 pandemic that has led to strict social measures in many states. Thus, it is the goal of this research to study the impact that the outbreak and its mitigation measures have had on the PLWHA. Specifically, we sought to highlight their concerns from sentiments expressed on social media based on posted tweets. By combining machine learning (ML) techniques such as textual mining and thematic analysis, we determined **14 major themes** as factors that are worth exploring. In this work, we originally extracted 2,839,091 tweets related to HIV/AIDS posted from March 2020 to April 2022. After initially doing data cleaning and preprocessing, we performed topic modeling using the Latent Dirichlet Allocation (LDA) topic model to extract 25 topics that are made up of 30 keywords each. The topics were then narrowed into 14 themes. The paper details the negative, positive, and neutral sentiment polarities which we highlight as concerning. These sentiments were determined using the **V**alence **A**ware **D**ictionary and s**E**ntiment **R**easoner (**VADER**) Sentiment Analysis Library with a 90% F1-score compared to TextBlob which showed a 53% F1-score. The research findings highlight issues affecting PLWHA during and post-pandemic such as high cost of medical care, late diagnosis of HIV, limited access to medications, stigmatization and victimization, absence of testing kits in hospitals, and lack of urgency in the development of vaccines or cure for HIV.

INDEX TERMS Sentiment analysis, thematic analysis, textual mining, tweets, HIV, AIDS, machine learning.

I. INTRODUCTION

HIV/AIDS continues to be an epidemic worldwide despite the significant volume of resources dedicated to fighting against its devastating effects. According to the most recent data made available by the World Health Organization (WHO), 38.4 million people were living with HIV/AIDS (PLWHA) worldwide with 1.5 million new infections within the same

year in 2021 [21]. The management of HIV/AIDS has always been imbalanced between rich and poor countries. For instance, before the outbreak of the coronavirus disease 2019 (COVID-19) pandemic, several countries were not able to cope with the huge demand associated with the management of HIV/AIDS [3], [22]. As a result, the world is dealing with a *syndemic* – which is the aggregation of two or more epidemics synergistically combining to cause an increased burden of disease [22]. It has also been noted by [23] that COVID-19 has shifted attention from HIV/AIDS thereby reducing the quality of care for PLWHA globally.

The associate editor coordinating the review of this manuscript and approving it for publication was Chien-Ming Chen.

Over the years, PLWHA have experienced undesired treatments. Some of the setbacks experienced include stigmatization, victimization, depression, physical health decline, etc. [9], [19], [20], [24], [40]. However, with the advent of the syndemic, it is important to study how PLWHA has been impacted. Also, it is important to study whether progress has been made to address the initially highlighted setbacks or whether the setbacks have been exacerbated by the syndemic.

In this regard, this research sought to investigate the concerns of the PLWHA by analyzing their opinions expressed on social media. According to [25] and [41], social media platforms encourage all-inclusive participation and promote health-related discourse, opinion, and experience sharing. Due to the anonymity (pseudo name) feature of some of these platforms, PLWHA can be expressive without being overly concerned about victimization [41]. At the same time, there is advancement in social media data analytics using machine learning (ML) [11]. With the use of the appropriate ML-based classifiers, the sentiment polarity (i.e., positive, or negative) can be detected from a posted message [19], [25], [28]. The research by [13] seeks to understand the evolution and emergence of research fronts about the HIV/AIDS virus using methods such as text mining and data analysis. As well, [10] discusses the increasing challenges and requirements for the analysis of large quantities of information in humanitarian health crises provided by different sources such as social media and the fundamental use of artificial intelligence techniques to analyze and segregate information based on its relevancy rapidly.

Moreover, medical professionals and computer science personnel are utilizing Machine Learning (ML) and Artificial Intelligence (AI) as powerful tools to target specific people at high risk of contracting HIV [3]. The reason it is so important to find these individuals is that they would make great candidates for Pre-exposure prophylaxis (*PrEP*), which is a drug that can be taken once a day to greatly lower the chance of an individual catching HIV. The most promising area that ML and AI are being used is social media because of how often people are using the apps for daily use. In the work of [3] for example, the researchers found that by using specific machine learning instruments they were able to predict 74% of condomless anal sex before it occurs on any given day. By having this integral information ''Just-In-Time'' texts could be sent to these individuals to try to eliminate these dangerous situations and therefore slow down the spread of this terrible virus. Further, In the case of HIV/AIDS specifically, ML and AI could be used to track the spread, enabling the prevention of further contamination as well as early treatment of those who have contracted it [2], [37].

Also, [29] used ML and thematic analysis to show the hidden themes of COVID-19. A corpus of COVID-related articles and publications was collected, and textual mining was used to gather common word clusters between the articles and group those articles together that contained relevant keywords and information. This text-mining process aids to identify needed topics quickly. Further, [33] suggests that deep learning models used iteratively can be a way to identify and promote new and changing HIV-related influencers online.

Moreover, [43] proposes an approach that will make it easy for people to analyze their health problems. Their work uses Twitter [44] to gather opinions as it is a place where people openly share their feelings. Almost 65% of internet users in India use Twitter and Facebook for health-related information [43]. This leads to the possibility of various public health issues, such as epidemics and diseases, being tracked using Twitter. This can aid in the early detection of an epidemic or other health issues. The results show a comparison of various phrases over *WordCloud*, concluding that the average accuracy of association between HIV-related illnesses is 98%. A recommended improvement of the study is to analyze the sentiment phrases, which will enhance the result and assess the sure, unsure, positive, negative, and neutral sentiments. Without sentiment analysis, similar terms with different sentiments could be incorrectly classified. Besides, [8] suggested after performing ML techniques on textual data that when designing interventions, attention should be paid to a particular demographic (older vs. younger) to help PLWHA achieve a better quality of life. This view is also shared by [15].

In our research, we posit that the combination of machine learning, thematic analysis, and sentiment analysis of tweets can reveal the frustrations of PLWHA during this time of the syndemic. Specifically, we sought to highlight their concerns from sentiments expressed on Twitter. Using the combination of machine learning (ML) techniques such as textual mining and thematic analysis, we determined **14 major themes** as factors that are worth exploring.

The 14 themes are (**a**) Sexuality and Lifestyle Choices of PLWHA, (**b**) Testing for HIV, (**c**) HIV/AIDS Prevention, (**d**) Healthcare and Support Resources, (**e**) Stigmatization, (**f**) HIV impact on minority groups such as Women and Children, (**g**) the Role of Government, (**h**) Related Diseases to HIV, (**i**) HIV Criminalization and Disclosure, (**j**) Research on HIV, (**k**) Misinformation on HIV, (**l**) Outbreak of HIV, (**m**) HIV Treatment, and (**n**) COVID-19 syndemic.

In this work, we originally extracted 2,839,091 tweets related to HIV/AIDS posted from March 2020 to April 2022. After initially doing data cleaning and preprocessing, we performed topic modeling using the Latent Dirichlet Allocation (LDA) topic model to extract 25 topics that are made of 30 keywords each. The topics were then narrowed into 14 themes. The paper details the negative, positive, and neutral sentiment polarities which we highlight as concerning. These sentiments were determined using the **V**alence **A**ware **D**ictionary and s**E**ntiment **R**easoner (**VADER**) Sentiment Analysis Library with a 90% F1-score compared to TextBlob which showed a 53% F1-score.

The research findings highlight the importance of understanding the issues affecting PLWHA during and post-pandemic such as late diagnosis of HIV, limited access to medications, stigmatization and victimization, absence of testing kits in hospitals, and lack of urgency in the development of vaccines or cures for HIV.

Overall, the work made the following contributions to fighting HIV/AIDS, machine learning for social good, and services engineering research:

- Researched and highlighted the frustrations of PLWHA during the syndemic – which is the combination of the COVID-19 pandemic and the HIV epidemic.
- Proposed fourteen (14) themes that are worth considering based on the outcome of the designed machine learning, thematic analysis, topic modeling, and sentiment analysis.
- The identified themes can inform policy formulation and governance decision-making on PLWHA and the global fight against HIV/AIDS.
- Developed a sentiment polarity classifier that can be used by researchers and individuals to group a tweet into a particular theme.

The remaining sections of the paper are arranged as follows. Section II features ongoing works in textual mining, social media data analysis, and machine learning specific to social good. Section III describes the proposed machine learning models of our tweets while the results of the models are discussed in Section IV. A full discussion of the results with recommendations is presented in Section V and the paper concludes in Section VI with our contributions and future research direction.

II. BACKGROUND WORKS

A. SOCIAL MEDIA, MACHINE LEARNING, AND TEXTUAL DATA MINING

There is continuous research on ways of integrating AI into workflows to enhance social policy interventions. In Table 1, we summarized some general areas where AI and ML are applied to social media data mining to address social problems.

According to [4], textual mining and data visualization can be used to determine how terminology such as "HIV infected" and "AIDS carrier" impacted the stigmatization of those infected with the disease. The main problem is that some terms such as "AIDS carrier" create an embarrassing and negative stigma around PLWHA which can affect their mental health and make their lives more difficult. Being professional when talking about the disease is the best way to decrease stigmatization around it and these practices should be carried on. The work curated textual data, converted it into a comma-separated values (CVS) format after cleaning, imported the data into a Microsoft SQL Server database, and then data visualization was used to identify trends. The data represents decreased stigmatization of those living with HIV when proper terminology is used to describe their condition.

Similarly, the research performed by [5] explored the use of textual mining of social media messages and visualization to identify health risk behaviors about HIV/AIDS.

In [1], the paper reviews the concept of AIDS as a social construction in the Italian context. The problem that motivated the design of the research was to identify the perception of people toward AIDS. Most of the representations of AIDS include moral values, human fears, and collective myths. The print media such as newspaper articles has played a massive role in communicating and educating society about AIDS misconceptions and ways of contracting, preventing, and controlling infections. This study analyzed text data from the newspapers for the period from 1985 to 1990 and 2005 to 2010 using text mining and thematic analysis. These processes were used to explore and analyze the data constituting unstructured text data which is enhanced through software that recognizes patterns, keywords, topics, concepts, and other data attributes. The results obtained indicated five thematic domains, which included emotional components, ethics and morality, law and order, medicine and biology, and institutional policy.

Furthermore, [6] employed textual mining and visualization techniques to analyze discourse on Twitter about pre-exposure HIV treatment, or Pre-Exposure Prophylaxis (PrEP). They used textual mining tools to cipher through the accumulated data for keywords such as Truvada, HIV, PrEP, pill, aids, pre-exposure, and a combination of those words as well. The tweets were then stored in a database, where they are stripped and analyzed for sentiment and other topics. Overall, the tool was successful in detecting and discovering what people were talking about and how they feel about PrEP. The analytics established that the sentiment was overall positive. The work is similar to [18] who observed that the word "PrEP" increased in usage over time as the term "pill" decreased on Twitter. This suggests that users are becoming more aware of the drug. But [38] concluded that most online posts are blockers to PrEP, halting the spread of information.

According to [7], universal HIV screening programs are costly, labor-intensive, and often fail to identify high-risk individuals. Therefore, the authors presented a model to improve HIV diagnosis by using natural language processing (NLP). To understand whether information found in clinical records can identify individuals at elevated risk of HIV infection, three machine-learning approaches that feature NLP for HIV risk were developed using machine-learning algorithms: (1) a baseline model with only structured EHR data, (2) baseline plus NLP topics, and (3) baseline plus NLP clinical keywords. The findings suggest that clinical notes exist as a valuable source of information on HIV risk factors including drug use and high-risk sexual activity and extracting variables using natural language processing improved the performance of predictive models for HIV risk. Moreover, [12] applied machine learning techniques on past HIV care data and a Bayesian network to predict how to care for PLWHA. The study can also help optimize health service resources for current and future viruses. Likewise, [26] opined that machine learning–based predictive analytics (MLPA) products could positively impact the prevention and treatment of HIV and AIDS, and the ability to change their treatment.

Moreover, the research by [14] shows how textual mining can be an effective strategy for recovering interactions and identifying specific proteins related to HIV. The research was able to show results that proved that textual mining was not

TABLE 1. Social media and machine learning.

Authors	Aim/Focus	Tools	Future Works
[45] Ye, T., Johnson, R., Fu, S., et al.	Utilizing machine learning to prioritize outreach and assistance to New York City tenants who face landlord harassment.	Gradient Boosting worked best when implemented with scikit-learn.	Field testing and more efficient methods of using the model are needed. Selection bias also needs to be reduced, as labels of buildings are only taken from NYC's Tenant Support Unit.
[46] Lakkaraju, H., Aguiar, E., Shan, C., et al.	Using machine learning to identify students at risk of not graduating high school and to provide intervention.	Random Forest consistently outperformed other models when tested against them using two school districts.	The algorithm can hypothetically be used for more than finding students at risk of not graduating on time, and it is suggested that it be used for things such as not applying to college or undermatching.
[47] Berberich, N., Nishida, T., & Suzuki, S.	Discussing the integration of concepts of harmony and tactful behavior into the ethics of artificial intelligence and its use of it.	Analysis of harmony across various cultures throughout history is used to argue for its importance.	Adding the concept of harmony into the core ethical principles of AI to lead the way towards stronger inclusion of Eastern perspectives in a Western-centric field.
[48] Islam, M., Kabir, M. A., Ahmed, A., et al.	Using machine learning to analyze social networks such as Facebook to scan for signs of depression.	Decision Tree shows the highest accuracy compared to other models. Linguistic Inquiry and Word Count were used to extract information.	The authors plan on utilizing other techniques for gathering data, as well as using more datasets to verify the efficiency and effectiveness of their techniques.
[49] Chavan, V. S., & Shylaja, S. S.	Devising methods of detecting cyberbullying using machine learning.	Natural language processing is used to process comments into usable forms, then N-gram, counting, and TF-IDF scores are used to construct feature vectors.	Future work can focus on detecting sarcastic comments.
[50] Lillywhite, A., & Wolbring, G.	Determining to what extent AI/ML engaged with disabled people.	Their algorithm has determined that disabled people are only covered in a limited capacity on Twitter.	Future research should include interviewing disabled people themselves, AI/ML policymakers, and funders.
[51] Alotaibi, S., Mehmood, R., Katib, I., et al.	Using machine learning to detect diseases using Twitter.	Sehaa gathered information to fill graphs with disease information in Saudi Arabia based on filtered and unfiltered tweets.	The functionality and scalability of the work could be improved.
[52] Braithwaite, S. R., Giraud-Carrier, et al.	Validating the use of machine learning algorithms against suicidality.	The DSI-SS screens for suicidal symptoms and assesses suicidality. LIWC extracts information from the texts.	Future research could explore tolerance for specific messaging approaches and inform effective health communication strategies.
[53] Oyebode, O., & Orji, R.	Using machine learning to detect diabetes factors in Nigeria using social media.	Their algorithm determined that the diabetes rate in Nigeria is climbing concerningly quickly.	Future work could develop a diabetes intervention application.
[54] Sravanthi, T., Hema, V., Reddy, S. T., et al.	Using machine learning to detect mentally distressed social media profiles.	NLP was used to determine what suicidal tendencies look like based on social media posts.	Expand on the stressors of Twitter itself in research.

only efficient at identifying instances where HIV-related data was discussed but was also able to identify even more pieces of data that are not currently listed in the human protein interaction database (HPID). While text mining is not able to replace the accuracy of manual data extraction, it can generate usable data much faster, and at a high success rate.

Likewise, [16] utilized textual mining to extract HIV data (topics and terms) from pathology reports. Using this data, evaluations based on risks and trends of the disease were compared to HIV patients in ten years, to record symptoms and other trends. [17] went a step further by studying the degree of support provided for PLWHA who subscribe to
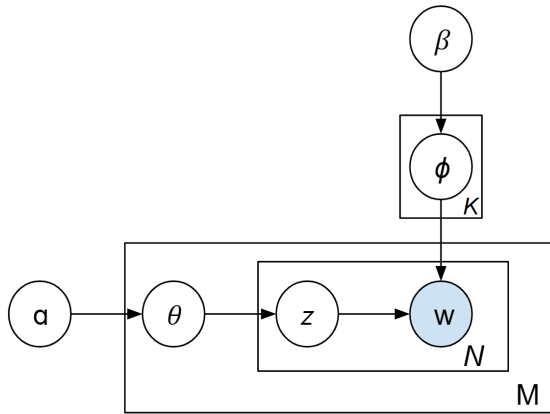
FIGURE 1. Graphical representation of LDA model.

TABLE 2. Summary of variables used in LDA.

Variable	Description
M	Number of documents
N	Number of unique words
w	Words in document
z	Topic assigned to words
K	Number of topics
Φ_k	Word distribution in topic k
θ_m	Topic distribution in document
α	Dirichlet prior for θ
β	Dirichlet prior for Φ

online forums. Additionally, [34] and [39] investigated the driving mechanisms of building interaction among people living with HIV/AIDS, particularly the largest online HIV communities which can lead to a better distribution of resources.

Moreover, text mining and visualization of social media data were used to determine the characteristics of the hidden population of PLWHA [19]. Findings from the study suggest that PLWHA tend to engage in conversations regarding their status and possible ways to improve their lives. The researchers utilized ML and NLP techniques to construct a network that helped them determine the relationship between desired variables in their study.

Also, [27] engaged AI and textual mining to see if they could identify the most relevant characteristics found in PLWHA that lead to the development of dementia. The paper used multiple techniques to see which one was the most accurate including logistic regression, decision tree, K-Nearest Neighbor (KNN), neural network, and random forest. Out of these, the random forest was the most accurate across all the different ranges of features when compared to the other techniques. However, a similar test on Twitter data by [35] revealed that logistic regression was the most accurate and fastest processing model.

Besides, [30] employed machine learning and visualization to focus on the use of data and algorithms to look for relevant articles in electronic databases and reference lists. The identified and valuable aspects of care have been grouped into seven themes: good relationships between health care professionals and patients, HIV expertise, continuity of care, easy access to services, and quality, access to high information and support, and participation in effective coordinated treatment and care decisions between HIV professionals and other healthcare professionals. In [42], a health-related topic classification tool was developed with an accuracy ratio of at least 85% for positive HIV/AIDS Internet health information inquiry for intervention.

Also, [31] used search engine data to monitor and forecast AIDS. A machine learning method, such as artificial neural networks (ANNs), is used to forecast AIDS incidences and deaths. Search trend data related to AIDS from Baidu.com

were collected and used as input variables of ANNs, and officially reported actual AIDS incidences and deaths were used as the output variable. Based on the monthly time series data, the work concluded that under the three criteria, the ANN method can lead to satisfactory forecasting of AIDS incidences and deaths, regardless of the change in the number of search queries. The results were similar to the findings in [32].

Also, [36] seeks to understand discussions related to HIV on the Twitter social media platform. Over 160,000 tweets were collected and analyzed using data mining techniques. The results of the study found that in terms of frequency, most tweets were about users' personal experiences with HIV, and encouragements to stop stigmatization.

III. METHODOLOGY
A. LATENT DIRICHLET ALLOCATION (LDA)
In textual mining, topic models can uncover hidden structures in data collection. In our work, the Latent Dirichlet Allocation (LDA) [55] is used to detect topics in tweets. The LDA is a generative probabilistic model of a collection of documents that are random mixtures of topics, where each topic is characterized by a set of words. LDA uses unsupervised learning and an iterative approach to discover topics in a collection of documents.

The graphical representation of LDA is shown in Figure 1. The boxes represent replicates. The outer box represents documents, while the inner box represents the repeated choice of topics and words within a document. LDA has the advantage of discovering hidden topics in data but cannot capture the correlations between topics generated. An explanation of the LDA variables is shown in Table 2.

Given a collection of documents (corpus) and the number of topics to discover (K), the generative process for LDA is as follows and expressed in equation 1:

1) For each topic(K), $K \in \{1 \ldots k\}$, draw a multinomial distribution over words
2) For each document $\{1 \ldots d\}$,
   a. generate a multinomial distribution over topics.
   b. For each word in the document:

i. Assign a topic to the word
ii. Generate the word from the selected topic.

$$p\left(w, z, \theta, \phi | \alpha, \beta\right) = \prod_{k=1}^{K} p\left(\phi_k | \beta\right) \prod_{m=1}^{M} p\left(\theta_m | \alpha\right) \cdots$$
$$\prod_{n=1}^{N} p\left(z_{m,n} | \theta_m\right) p\left(w_{m,n} | \phi z_{m,n}\right) \quad (1)$$

where $p\left(w, z, \theta, \phi | \alpha, \beta\right)$ is the joint probability of a word in a document that belongs to topic k,

B. OVERALL WORKFLOW

The workflow of the proposed approach as shown in Figure 2, is divided into 5 stages: Data Collection, Data Preprocessing, Topic Modeling, Thematic Analysis, and Sentiment Analysis.

Initially, we extracted HIV/AIDS-related tweets from Twitter and preprocessed them to remove irrelevant data. Next, we identify the topics on HIV/AIDS using Latent Dirichlet Allocation (LDA) topic modeling and then categorized related topics into themes. Finally, using sentiment analysis, the positive and negative factors influencing each theme were detected. The modules of the proposed work are discussed as follows.

C. DATA COLLECTION

Related HIV/AIDS Tweets were collected through the Twitter API [56] based on the following keywords: "#HIV", "HIV", "HIV/AIDS", "#PLWHIV", and #PeoplewithHIV. We retrieved tweets between March 1, 2020 and April 30, 2022. Overall, 2,839,091 tweets were extracted and converted into a CSV format for text analysis.

In Figure 3a, we showed the visualization of the sources of the tweets globally, and Figure 3b presents a visualization of the cleaned/pre-processed tweets.

D. DATA PREPROCESSING

Using the Python programming language, Natural Language Processing (NLP) techniques were implemented to clean and transform the extracted tweets into a usable format for topic modeling and sentiment analysis. The following steps were engaged to perform data preprocessing.

1. *Text Cleanup*
   a. Removal of non-English tweets: Machine Translation models are likely to produce errors during translation which can cause the accuracy of sentiment classifiers to drop. Hence, in our study, we focus on tweets posted in the English Language only. The *FastText* library is used to filter out non-English tweets. Tweets written in a combination of English and other languages were also discarded.
   b. Removal of known spam accounts and bots: A list of known spam accounts is compared with the usernames of all accounts in the extracted tweets to discard any spam and bot accounts.
   c. Removal of irrelevant tweets: Tweets using our defined search keywords as hashtags but do not mention HIV/AIDS in the main sentence are removed. Examples are advertising tweets.
   d. HyperText Markup Language (HTML) entities are decoded by using the unescape() function. Accented characters are transformed into regular characters. We remove retweets (RT), mentions, URLs, and hashtags as they are irrelevant for analysis. The decoding was successful using the HTML library.
   e. Expand contractions and slang: We used the contractions library to expand contracted text. For example, I'll becomes I will. Also, a list of internet and text slang scraped from the *noSlang* website is used to replace slang in tweets with their meanings.
   f. Remove duplicate tweets: We filtered out duplicate tweets to obtain a unique dataset. After removing duplicates, we had a total of 214,262 tweets for further text analysis.
2. *Tokenization*: All punctuations used in the tweets are removed and then tokenized. Tokenization is the splitting of tweets into a list of words.
3. *Removal of stopwords*. A custom list of stopwords was created in addition to the NLTK's stopwords for the removal of stopwords. The list of custom stopwords used are: [*"https", "http", "hiv", "plwhiv", "plhiv", "human", "humans", "aids", "bitch", "f∗ck", "shit", "d", "ad" … and other expletives*]
4. *Lemmatization*: We transformed the words in the tweets into their root word. This is to create a dictionary of unique words for the set of tweets (corpus). Lemmatization is more effective than stemming since it converts the word into its root word rather than removing the suffices.

Stopwords and punctuations contribute to the overall polarity of texts in lexicon-based sentiment analysis [25], hence we maintained them during the sentiment analysis of the tweets. However, for the detection of topics, the data was converted into lowercase, and punctuations and stopwords were removed. We then lemmatized the tweets and converted them into a document word matrix for the LDA algorithm. Figure 4 shows the number of unprocessed and preprocessed tweets for each month.

E. DATA VECTORIZATION

The preprocessed tweets are converted into numerical feature vectors for the LDA. The Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) are the most widely used feature extraction techniques to encode text into numerical feature vectors for machine learning algorithms.

The BoW approach is selected to create a document word matrix input for the LDA. The reason is that LDA is a generative probabilistic model based on document and term count. TF-IDF focuses on the relevance of words using weights and will reduce the distribution of word frequencies.
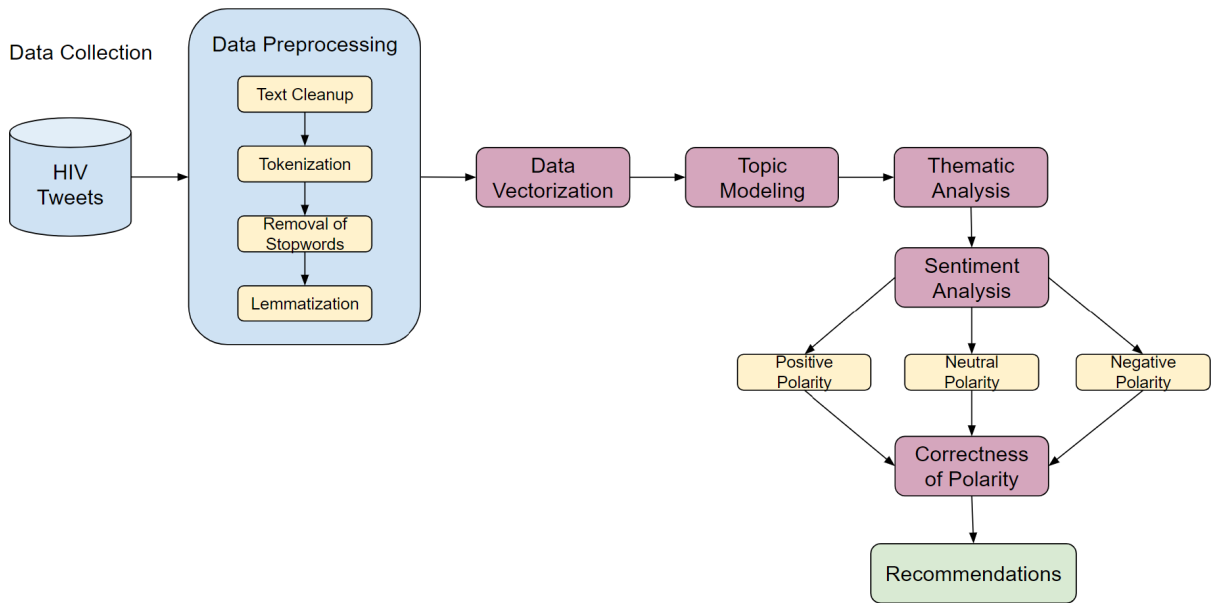
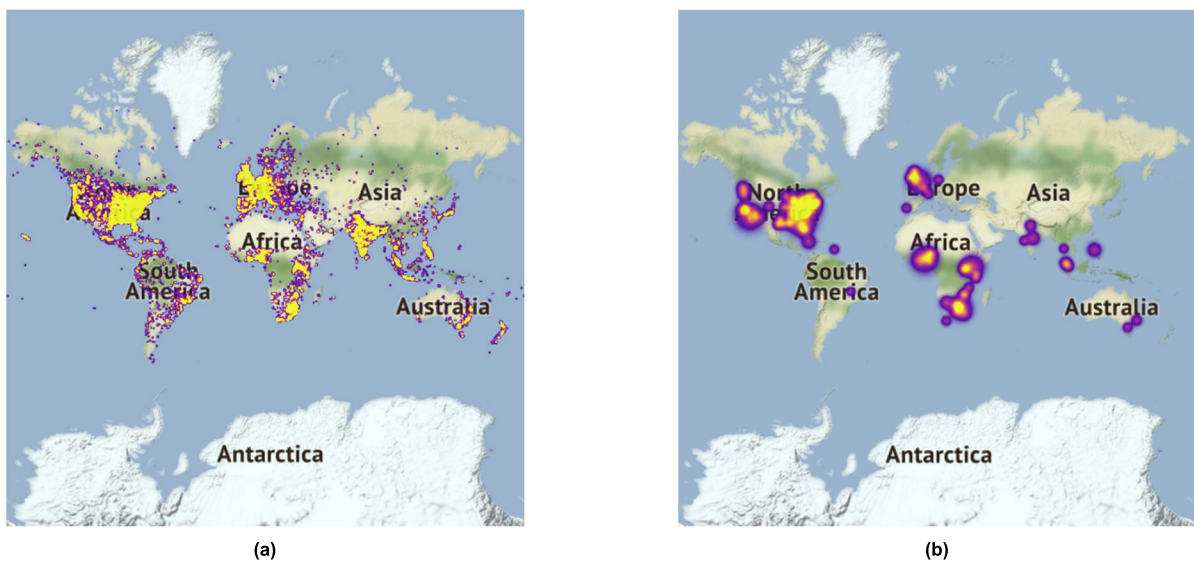**FIGURE 2.** Overview of the proposed approach.

FIGURE 3. (a) Original global tweet data sources. (b) Resultant tweets after data preprocessing.

Thus, fitting LDA with a TF-IDF vectorizer will result in less inter- or intra-document statistical structure according to [55].

The BoW vectorizes terms by counting the number of occurrences of each unique term. The *CountVectorizer* from the Sklearn feature extraction library was used to generate the document term matrix. Table 3 shows the BoW feature vectors of two sample tweets from our HIV dataset. Also, see some original and preprocessed tweets below.

Original T1: my viral load has reached undetectable status as of today which means my hiv is virtually untransmittable you all do not know what this means to me

Preprocessed T1: viral load reach undetectable status today mean virtually untransmittable mean

Original T2: recent check-up:cd: viral load (vl): (unde-tectable)my happiness and security? Priceless

Preprocessed T2: recent check viral load undetectable happiness security priceless

F. TOPIC MODELING

We implemented the LDA topic modeling algorithm using the Sklearn package. To select the optimal number of topics for our dataset, we trained LDA model with different number of topics (k), $k \in \{10:50\}$ where k starts at 10 and ends at 50.
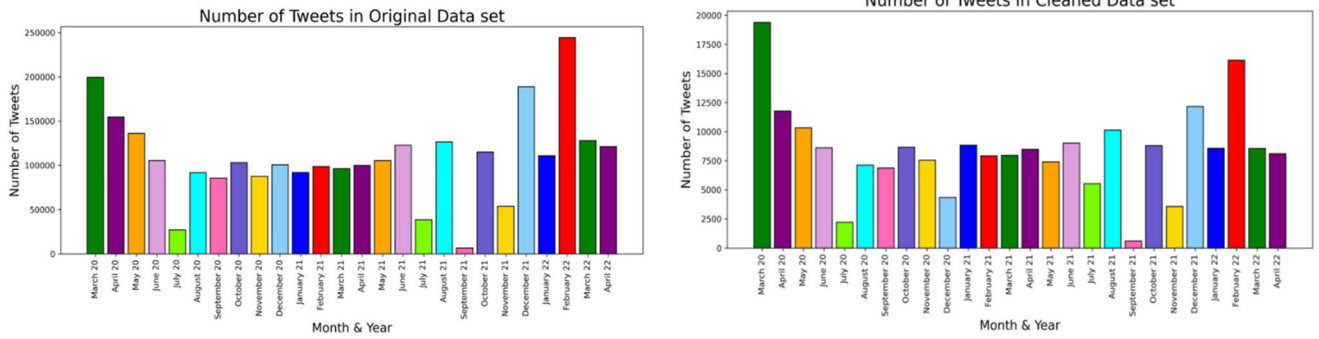
FIGURE 4. The number of tweets in original and cleaned dataset.

TABLE 3. Bag of words features on preprocessed tweet.

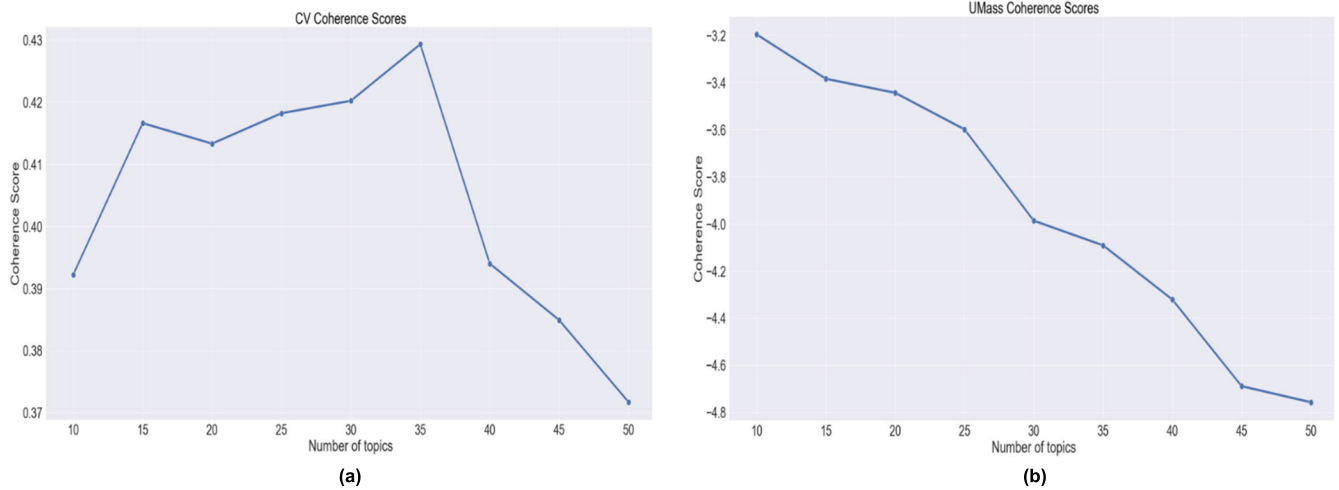| T	check	happiness	load	mean	priceless	reach	recent	security	status	today	undetectable	untransmittable	viral	virtually
1	0	0	1	2	0	1	0	0	1	1	1	1	1	1
2	1	1	1	0	1	0	1	1	0	0	1	0	1	0

FIGURE 5. (a) CV and (b) UMass coherence score of the topics.

The LDA model analyzes the distribution over topics and extracts the document term matrix (BoW) into topics by computing the weighted frequency of the terms each topic can have. The topic coherence score is used to decide the number of topics, where the topic coherence score assesses the quality of topics generated by measuring the sematic similarity of words in the topic. The coherence score is computed using the tmtoolkit [57] since the Sklearn package does not have the implementation of the coherence score.

Two types of coherence metrics were considered to select the best number of topics: CV and UMass coherence score. A model with coherent topics will have a higher score. In the case of the UMass coherence metric, the number of topics with a score closer to zero indicates a higher topic coherence. As shown in Table 4 and Figure 5, the CV coherence metric indicates the optimal number of topics varies between 15 and 35, while the UMass coherence metric indicates the best number of topics ranges from 10 to 30. We opted to train the LDA model with 25 topics captured on HIV as discussed over time. As illustrated in Figure 5, selecting a topic number above 30 might cause the model to overfit.

G. THEMATIC FORMATION

Thematic formation (analysis) is conducted to categorize topics into themes to discover insights from the HIV tweets. We manually analyzed the top cluster of words generated by the LDA model to interpret and label topics. The labeled topics were reviewed by experts to ensure the topics were related to HIV/AIDS.
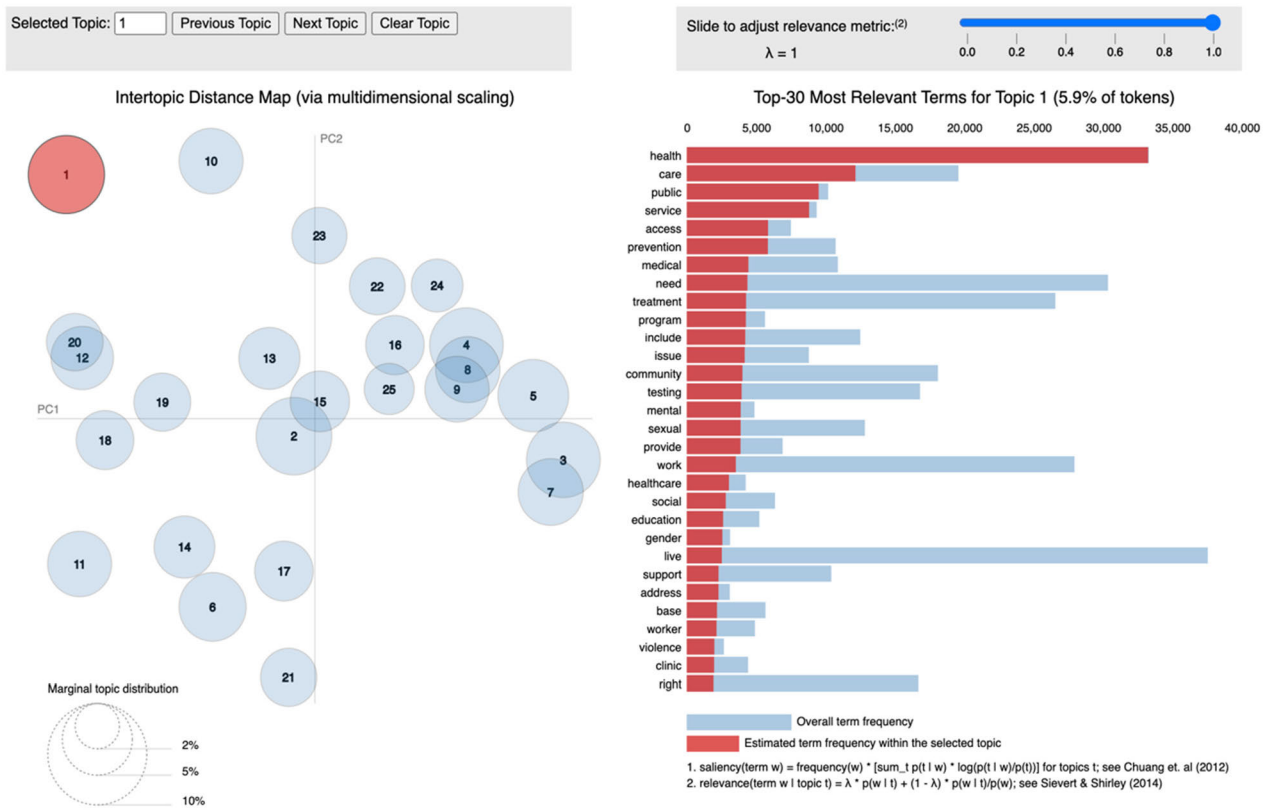
FIGURE 6. pyLDAvis visualization of 25 topics.

TABLE 4. Number of topics with their coherence score.

Number of	Coherence Score	
Topics (k)	CV	UMass
10	0.3922	-3.1958
15	0.4166	-3.3843
20	0.4133	-3.4439
25	0.4182	-3.5991
30	0.4202	-3.987
35	0.4293	-4.0915
40	0.394	-4.3218
45	0.3849	-4.6896
50	0.3717	-4.7578

TABLE 5. Sentiment score for assigning sentiment to tweets.

Sentiment	VADER	TextBlob
Positive	>= 0.05	>0
Neutral	> -0.05 - < 0.05	0
Negative	<=-0.05	<0

Topics that are related are combined to form a theme. We generated 14 themes from the 25 topics.

H. SENTIMENT ANALYSIS

We analyzed sentiments in the tweets for each theme using two well-known lexicon-based sentiment classifiers: Valence Aware Dictionary and sEntiment Reasoner (VADER) [58] and TextBlob. Lexicon-based sentiment classifiers use a set of pre-defined lexicons to classify a text as positive, neutral, or negative based on the polarity score of the text. The data

is classified as positive, neutral, or negative based on the polarity score of the tweets. TextBlob is a textual data processing library based on Natural Language Tool Kit (NLTK) and pattern.

Meanwhile, VADER is a lexicon and rule-based sentiment analysis tool for analyzing social media texts. For a given piece of text, lexicon-based classifiers analyze the text to verify if words in the text are present in the lexicon database and assign a lexicon rating to each word found. The polarity score is computed by normalizing the sum of all the lexicon ratings to the range of -1 and 1. The polarity score is a float value between -1 and 1 where tweets with a polarity score greater than zero are assigned positive sentiments, equal to zero are neutral sentiments and less than zero are negative sentiments.

Since VADER was designed specifically for social media texts, it can comprehend and analyze sentiments expressed in social media slang better than TextBlob. It has limitations

TABLE 6. Topics and relevant terms generated by unsupervised LDA model from HIV tweets.

Topic ID	Relevant Terms	Topic Label	Number of Tweets (%)
1	health, care, public, service, access, prevention, medical, need, treatment, program	Healthcare	11,148 (5.20%)
2	Year, life, live, world, pandemic, save, crisis, change, die, epidemic, learn	Pandemic year	13,051 (6.09%)
3	Tell, say, get, kill, die, know, lie, right, wrong, keep	Information	15,836 (7.39%)
4	Test, get, positive, result, negative, free, check, blood, kit, home	Testing	15,875 (7.41)
5	Friend, remember, back, need, time, family, leave, lose, positive, brother	Relatives	12,783 (5.97%)
6	Virus, COVID, Variant, Discover, Immunity, Scientist, Spike, Mutation, Genetic, Antibody	Covid-19	10,032 (4.68%)
7	Play, catch, herpe, covid, stds, real, magic, scared, vaccinate, fake	Sexually Transmitted Diseases	12,652 (5.90%)
8	Post, good, question, tweet, medium, video, take, watch, social, news	Social Media	9,621 (4.49%)
9	Stigma, Love, Feel, Educate, Ignorant, Need, Dirty, Clean, Mean, Comment	Stigmatization	9,868 (4.61%)
10	Support, live, awareness, help, stigma, community, fight, discrimination, advocate, member	Create Awareness	8,272 (3.86%)
11	Immune, Infection, Death, System, Virus, Disease, Pandemic, Symptom, Epidemic, Attack	Immunity	7,844 (3.66%)
12	Drug, country, work, company, global, fund, fight, research, government, development	Drug Research	7,450 (3.48%)
13	Medication, prevent, prep, healthy, exposure, live, pill, life, treatment, prophylaxis	Pre-exposure prophylaxis	6,569 (3.07%)
14	Covid, disease, spread, rate, compare, high, country, infectious, contact, vaccination	Outbreak of HIV	7,128 (3.33%)
15	Gay, Black, Community, Commercial, Straight, Queer, white, donate, blood, ban	Sexuality and Lifestyle choices	8,138 (3.80%)
16	Woman, child, mother, rape, positive, pregnant, birth, infect, fear, family	Women and Children	6,406 (2.99%)
17	Vaccine, Covid, Year, Trial, Develop, Find, Decade, MRNA, Clinical, Phase	Covid vaccine	7,908 (3.69%)
18	Blood, risk, high, heart, disease, condition, transplant, pressure, disorder, diabetes	Related Diseases	5,593 (2.61%)
19	Read, cell, write, history, pandemic, research, epidemic, study, work, plague	Research on Diseases	5,579 (2.60%)
20	Treatment, viral, patient, undetectable, load, infection, transmit, therapy, antiretroviral, diagnose	HIV Treatment	5,310 (2.48%)
21	Cure, cancer, death, cold, disease, suicide, virus, heart, diabetes, malaria	Related Diseases	6,820 (3.18%)
22	Sex, safe, unprotected, partner, shame, practice, teach, education, risk, contract	Forms of HIV Transmission	5,647 (2.64%)
23	Positive, status, person, know, partner, disclose, law, arrest, relationship, state	HIV Criminalization and Disclosure	5,846 (2.73%)
24	Condom, Wear, Mask, PREP, Protect, Safe, Partner, Trust, Risk, Contract	Protection	5,255 (2.45%)
25	Government, refuse, pay, money, state, outbreak, medicine, danger, stigmatize, federal	Role of Government	3,631 (1.69%)

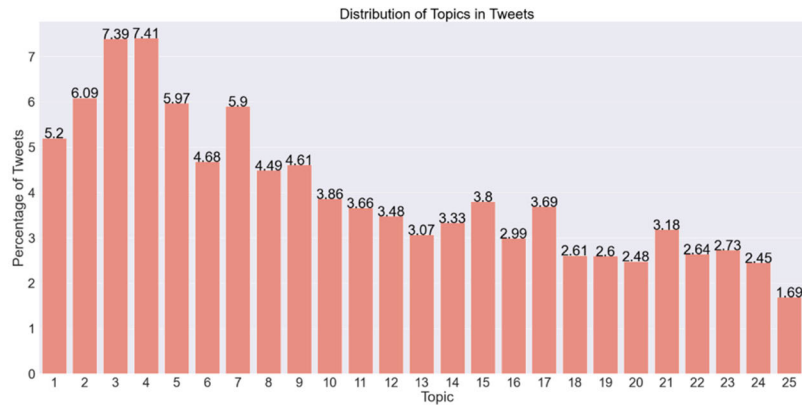**FIGURE 7.** Word cloud of (a) Topics 1 and (b) Topics 4.

FIGURE 8. Percentage of topics in tweets.

such as misinterpreting sarcasm, language-specific, and classification limited to the set of pre-defined lexicons.

For the VADER classification of sentiments, we allocated positive sentiment to tweets with a compound score greater than or equal to 0.05, negative sentiments to tweets with a compound score of less than or equal to −0.05, and neutral sentiments to tweets with a compound score within the range of −0.05 and 0.05. In the case of TexBlob, the polarity score to assign sentiments is between −1 and 1. These score allocations used to compute the sentiments are shown in Table 5.

IV. RESULTS

The experiments of this study were implemented in Python and run on Google Colab notebooks. We leveraged Python built-in libraries and NLTK to preprocess the tweets for analysis. We used the LDA library from the Sklearn package to extract topics.

A. TOPICS

We selected 25 as the optimal number of topics based on the computed UMass coherence scores discussed in Section III to discover representative topics in the collected HIV/AIDS tweets. The pyLDAvis [59] package in Python is used to visualize and interpret topics generated by the LDA topic model. As shown in Figure 6, each bubble represents a topic. The size of the bubble indicates the predominance of the topic in the corpus. A larger bubble indicates a higher number of tweets on a topic. The blue and red bars represent the overall term frequency of words and estimated term frequency within the selected topics respectively. The relevant term with the

longest red bar is the most frequent word used in a selected topic. Table 6 shows the top 10 relevant terms for each topic. We manually assigned labels to each topic using the relevant terms with a higher weight.

Furthermore, to confirm the assigned topic labels, we generated the word clouds of tweets for all topics. As illustrated in Figure 7, the bigger and bolder letters highlight the important words in Topic 1 (health, HIV, care, medication, and public) and Topic 4 (HIV, test, results, and kits). The document-by-topic matrix is used to assign topics to each tweet, the dominant topic for a tweet has a higher weight. We manually validated the assignment of topics for tweets by reviewing randomly sampled tweets from each topic.

As shown in Table 6, the top six predominant topics in our HIV tweet dataset are Topic 4 (Testing), Topic 3 (Information), Topic 2 (Pandemic year), Topic 5 (Relatives), Topic 7 (Sexually Transmitted Diseases), and Topic 1 (Healthcare). Figure 8 shows the percentage of tweets for each topic. Topic 25, representing the role of the Government on HIV, has the lowest percentage of tweets at a value of 1.69%. Topic 4 which highlights the concerns of HIV testing is the most prevalent discussion among users with 7.41% of tweets. We set no threshold for a topic ratio of tweets to filter out insignificant topics because we wanted to study and track all topics generated by the model over time.

B. COMBINING TOPICS INTO THEMES

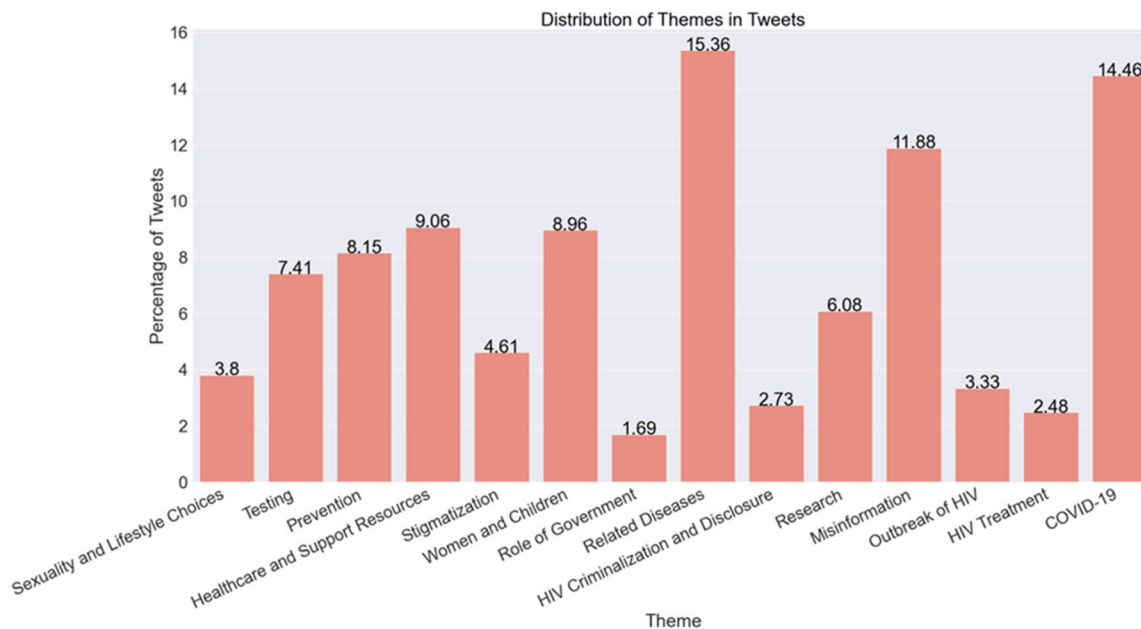The 25 topics were grouped into 14 main themes for interpretation and analysis to gain insights from the HIV/AIDS

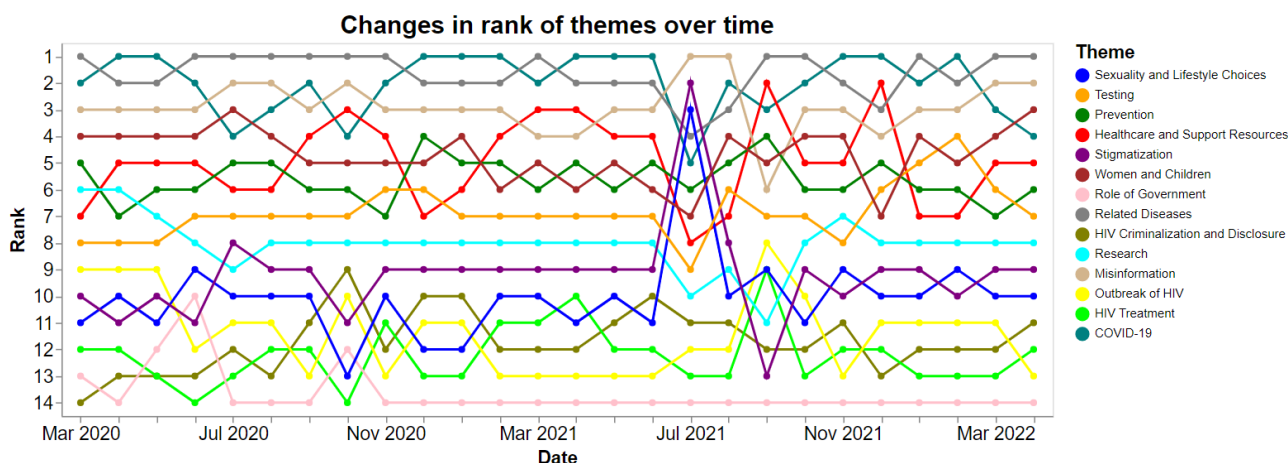**FIGURE 9.** Percentage of themes in tweets.

FIGURE 10. The rank of themes over time.

tweets. The word cloud for each theme is illustrated in Appendix. Related topics were merged into a theme. For instance, the Prevention theme is the combination of Topic 13, 22, and 24 which represents Pre-exposure prophylaxis, Forms of HIV Transmission, and Protection respectively. Figure 8 shows the percentage of tweets for each theme. The top 6 dominant themes in the HIV/AIDS corpus are Related diseases, COVID-19, Misinformation, Healthcare Support and Resources, Women and Children, and Prevention. The distribution of HIV Treatment, HIV Criminalization, and Disclosure, and the Role of Government themes were below 3%.

The prominence of themes throughout our collected tweets is illustrated in Figure 9. The top four most prevalent themes

focused on Related diseases, COVID-19, Misinformation, and Women and Children from March 2020 to June 2020 and March 2022 to April 2022. The Role of Government and HIV Criminalization and Disclosure themes were infrequently discussed on Twitter in the first two months.

Also, the changes in the rank of themes over time are plotted in Figure 10. The Role of Government was the lowest-ranked theme from November 2020 to April 2022. The testing theme remained in the mid-positions from May 2020 to June 2021. Similarly, the Research theme ranked 8th place, between the period of August 2020 to June 2021 and the last five months.

Stigmatization ranked 9th place from November 2020 to June 2020, then peaked at 2nd position in July 2021.

TABLE 7. Performance results of lexicon-based classifiers on themes.

Theme	VADER			TextBlob		
	Precision (%)	Recall (%)	F1-Score (%)	Precision (%)	Recall (%)	F1-Score (%)
Sexuality and Lifestyle Choices	96	95	95	58	50	49
Testing	87	85	85	69	62	64
Prevention	92	92	91	63	60	59
Healthcare and Support Resources	80	79	79	51	52	51
Stigmatization	95	95	95	62	58	58
Women and Children	94	93	93	61	55	56
Role of Government	84	78	79	68	57	60
Related Diseases	92	92	92	61	50	50
HIV Criminalization and Disclosure	88	87	87	59	57	55
Research	96	96	96	55	51	51
Misinformation	88	82	82	64	56	56
Outbreak of HIV	79	69	72	57	40	42
HIV Treatment	80	78	79	48	49	48
COVID-19	97	97	97	55	50	50

TABLE 8. Performance results of lexicon-based classifiers on total sampled tweets.

Classifier	Sentiment	Precision (%)	Recall (%)	F1-Score (%)
VADER	Negative	95	86	91
	Neutral	95	87	91
	Positive	82	96	88
	Weighted Average	**91**	**90**	**90**
TextBlob	Negative	73	41	52
	Neutral	41	52	46
	Positive	48	71	57
	Weighted Average	**59**	**53**	**53**

However, it became the least prominent theme discussed in September 2021. Correspondingly, Sexuality and Lifestyle Choices were the 3rd dominant theme debated on Twitter. The ongoing exchange of views on Healthcare Support and Resources peaked in October 2020, September 2021, and December 2021.

Meanwhile, discourse on Sexuality and Lifestyle Choices and HIV Treatment were both ranked at the 9th position in September 2021. Prevention is the 5th theme that triggered the attention of the Twitter audience at the beginning of the period, then became the 4th dominant theme discussed in December 2020 and September 2021. In the first three months of the period (March 2020 to May 2020), the Outbreak of HIV was the 9th theme discussed on the social space, however, it was rarely discussed from February 2021 to June 2021 and the last month of the period (April 2022).

C. SENTIMENT ANALYSIS
We employed two lexicon-based sentiment classifiers, *VADER* and *TextBlob*, to identify the concerns positively or negatively influencing the themes identified. We manually annotated 5% of random tweets from each theme as the ground truth to validate the performance of the sentiment classifiers. In the context of an imbalanced dataset, precision, recall, and F1-score were used as evaluation metrics to measure the performance of the classifiers. The evaluation metrics are computed through the classification report of scikit-learn library. The evaluation metrics are calculated in Equations (2)-(4):

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

$$F1 - score = 2\left(\frac{Precision * Recall}{Precision * Recall}\right) \qquad (4)$$

where TP, FP, TN, and FN mean True Positive, False Positive, True Negative, and False Negative respectively. We evaluated the performance of VADER and TextBlob using weighted average precision, recall, and F1-score. The weighted average calculates the evaluation metrics for each sentiment class and

TABLE 9. Distribution of sentiments and their proportion in each theme.

| Theme	Topic Number	Positive Sentiment (%)	Negative Sentiment (%)	Neutral Sentiment (%)	Number of Tweets (%)
Sexuality and Lifestyle Choices	15	2,632 (32.3%)	3,270 (40.2%)	2,236 (27.5%)	8,138 (3.80%)
Testing	4	7,685 (48.4%)	4,224 (26.6%)	3966 (25%)	15,875 (7.41%)
Prevention	13, 22,24	8,486 (48.6%)	6,138 (35.1%)	2,847 (16.3%)	17,471 (8.15%)
Healthcare and Support Resources	1, 10	11,154 (57.4%)	5,362 (27.6%)	2904 (15%)	19,420 (9.06%)
Stigmatization	9	4,269 (43.3%)	4,063 (41.2%)	1536 (15.6%)	9,868 (4.61%)
Women and Children	5, 16	7,447 (38.8%)	8,808 (45.9%)	2,934 (15.3%)	19,189 (8.96%)
Role of Government	25	1,212 (33.4%)	1,940 (53.4%)	479 (13.2%)	3,631 (1.69%)
Related Diseases	7,11,18, 21	9,859 (30%)	17,101 (52%)	5949 (18.1%)	32,909 (15.36%)
HIV Criminalization and Disclosure	23	2,638 (45.1%)	2,349 (40.2%)	859 (14.7%)	5,846 (2.73%)
Research	12, 19	5,743 (44.1%)	4,528 (34.8%)	2758 (21.2%)	13,029 (6.08%)
Misinformation	3, 8	8,805 (34.6%)	12,582 (49.4%)	4,070 (16%)	25,457 (11.88%)
Outbreak of HIV	14	2,215 (31.1%)	3,417 (47.9%)	1496 (21%)	7,128 (3.33%)
HIV Treatment	20	2,467 (46.5%)	1,609 (30.3%)	1234 (23.2%)	5,310 (2.48%)
COVID-19	2, 6, 17	12,391 (40%)	12,903 (41.6%)	5697 (18.4%)	30,991 (14.46%)
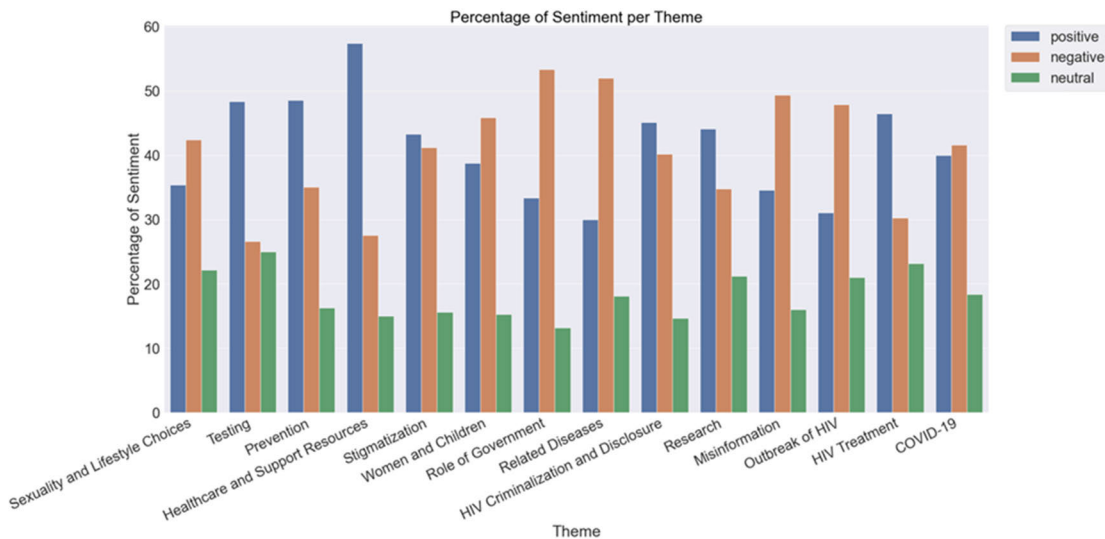
FIGURE 11. Percentage of sentiment per theme.

finds their average weighted by the number of true instances for each sentiment class. The weighted average for the evaluation metrics is defined in equation (5) as follows:

$$Weighted\ Average = \frac{1}{\sum_{l \in L} |y_l|} \sum_{l \in L} |y_l| \emptyset \left(y_l, \hat{y}_l \right) \quad (5)$$

where L is the set of sentiment classes, $\hat{y}$ is the predicted sentiment class, y is the true sentiment class, y_l is the subset of y with sentiment class l, and $\emptyset \left(y_l, \hat{y}_l \right)$ computes the precision, recall, or F1-score for the true and predicted sentiment classes that have the sentiment class l.

The weighted average precision, recall, and F1-score of VADER and TextBlob on each theme is reported in Table 7. VADER performed better than TexBlob on each theme. Randomly sampled tweets were merged to determine the overall performance of the lexicon-based sentiment classifiers. The performance of each classifier on overall sampled tweets is shown in Table 8.

In the case of VADER's performance, both negative and neutral sentiments have a precision score of 95% and an F1-score of 91%. The positive sentiment has a precision, recall, and F1-score of 82%, 96%, and 88% respectively.
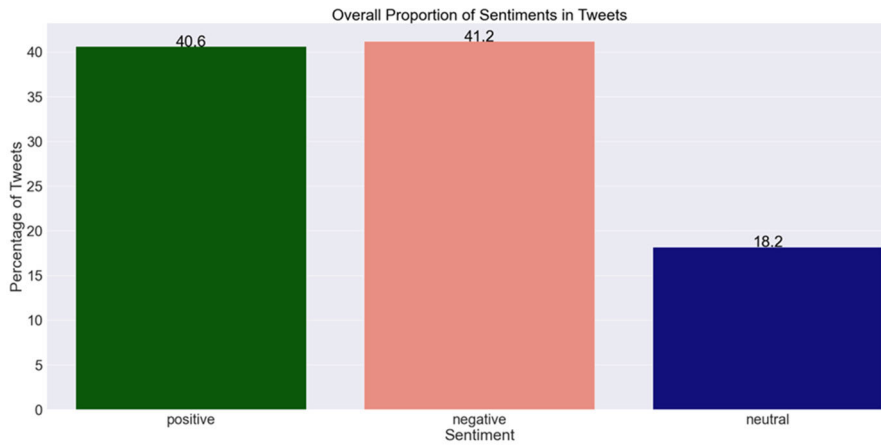
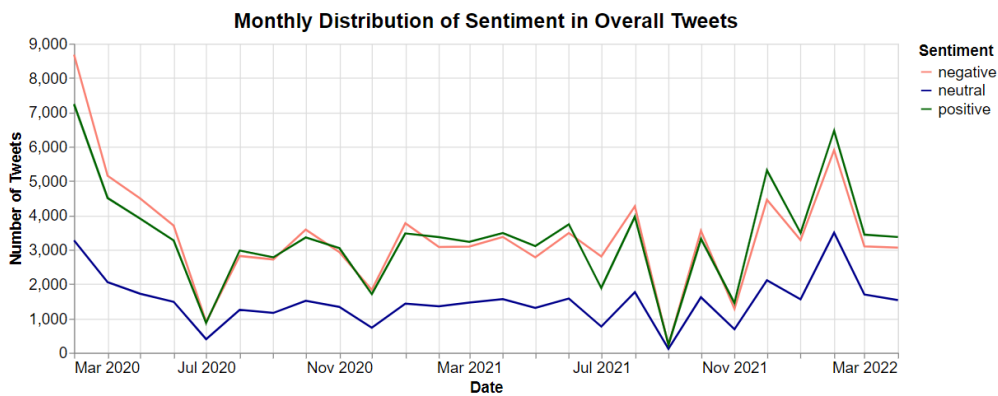**FIGURE 12.** Overall sentiment distribution in tweets.

FIGURE 13. Monthly trends of sentiment for overall tweets.

On the other hand, TextBlob achieved an F1-score of 57% for positive, 52% for negative, and 46% for neutral. Overall VADER outperformed TextBlob with a weighted average score of 91% for precision, and 90% for both recall and F1-score. TextBlob achieved a weighted average precision of 59%, and 53% for both weighted average recall and weighted average F1-score. The high performance of VADER is due to its ability to comprehend and analyze sentiments expressed in social media slang better than TextBlob. However, both tools have the limitation of not being able to detect sarcastic tweets. Based on this performance, we employed VADER to identify the sentiments of the remaining tweets in each theme.

In Table 9, the sentiments identified for each theme and their corresponding number of tweets are shown. The percentage of sentiment for each theme is shown in Figure 11. For the Sexuality and Lifestyle Choices tweets, most of the tweets were negative with a percentage of 40.2%, followed by positive tweets with a value of 32.3%. For the Stigmatization theme, the proportion of positive, and negative sentiment is 43.3%, and 41.2% respectively. Similarly, for Testing and Prevention themes, the percentage of positive sentiment is the highest, followed by negative sentiment. Meanwhile, in

the case of the Healthcare and Support Resources theme, more than 50% of the tweets had positive sentiment and the percentage of negative sentiment is 27.6%. However, for Women and Children, the Role of Government, and Related Diseases, negative tweets were dominant over positive tweets. The top sentiment in HIV Criminalization and Disclosure, and Research themes are positive followed by negative sentiments. For Misinformation, Outbreak of HIV, and COVID-19 themes, the distribution of negative sentiment was higher than positive sentiment. Nevertheless, in the HIV Treatment theme, a higher percentage of Twitter users expressed positive sentiments about the theme. Overall, we observed that neutral sentiment was the lowest for each theme.

The overall distribution of sentiments for the collected HIV/AIDS tweets is shown in Figure 12. The most prevalent sentiment was negative, that is 41.2%, followed by positive sentiment with a proportion of 40.6%, and neutral sentiment with the lowest percentage of 18.2%.

We analyze the trends of sentiments expressed on Twitter timelines throughout the tweets collection phase. Figure 13 highlights the changes in sentiment from March 2020 to April 2022 for the overall tweets. At the beginning of the
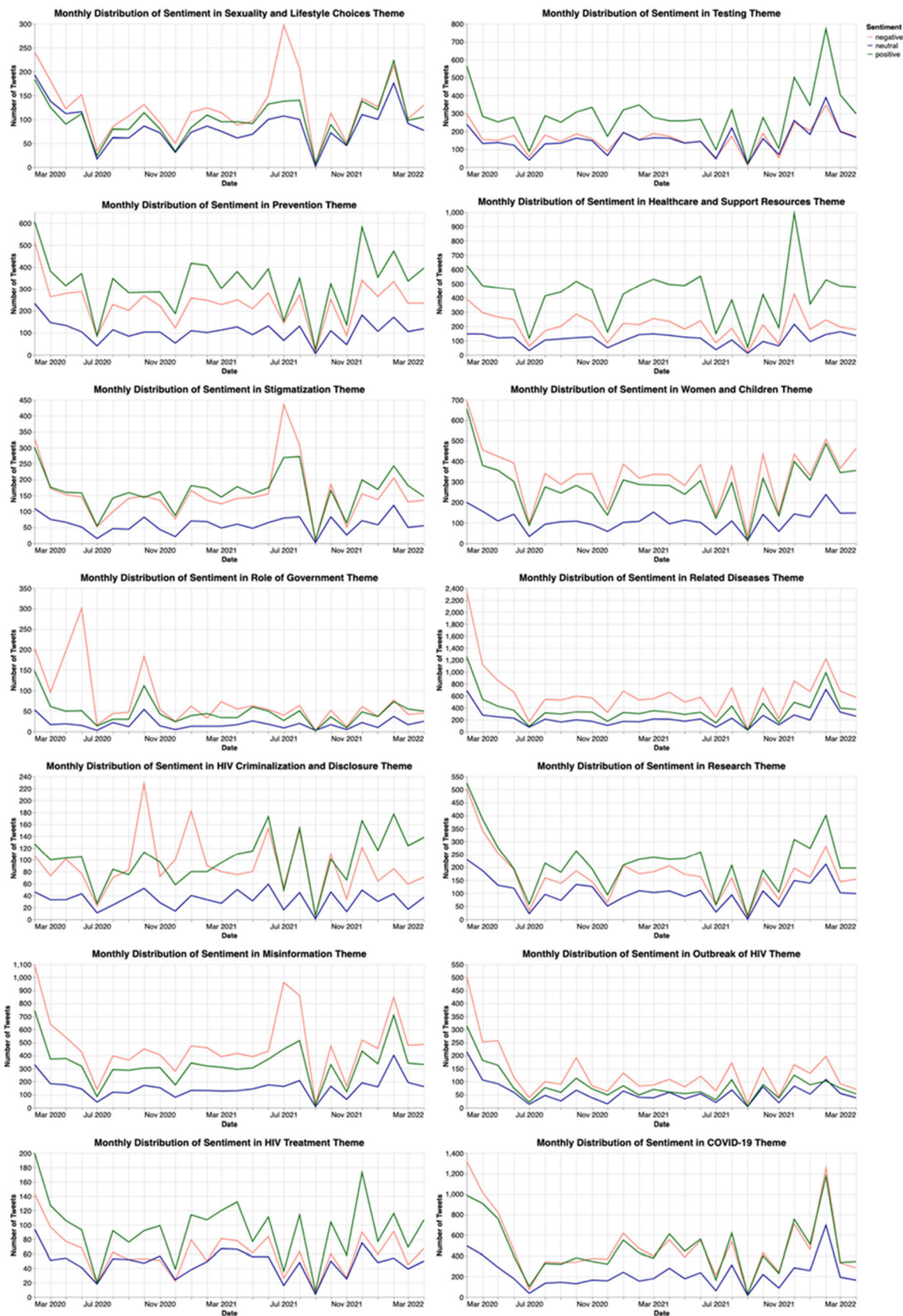
FIGURE 14. Sentiment trends over time for each theme.

period, there was a higher number of tweets for all sentiments, however, the dominant sentiment was negative. This indicates the increase in Twitter usage in the earlier period for the users in our collected tweets. In July 2020, the number of tweets for all sentiments decreased with positive and negative sentiments having a close value for the proportion of tweets. From August 2020 to August 2021, the number of tweets for both positive and negative sentiments was similar. In September 2021, the number of tweets for all sentiments was relatively low. However, from October 2021 to the end of the period, the number of tweets for all the sentiments grew steadily. At the end of the period, the most prevalent sentiment was positive, followed by negative, and lastly neutral sentiment.

The evolution of sentiment throughout the period for each theme is shown in Figure 14. Negative sentiments make up most tweets in the Sexuality and Lifestyle choices theme. However, in February 2022, the number of positive tweets was slightly higher than the number of negative tweets. A striking growth of negative sentiments was observed in July 2021.

For the Testing theme, user discussions throughout the period were positive. The proportion of positive sentiment increased sharply in February 2022. The negative and neutral sentiments had a close number of tweets from the beginning of 2021 to the end of the period.

However, the neutral sentiment was higher than the negative sentiment in August 2021, December 2021, and February 2022. Similarly, for Prevention, Healthcare, and Support Resources themes, the prevalent sentiment expressed throughout the period was positive. Positive discussions on Healthcare and Support Resources theme increased dramatically in December 2021.

Meanwhile, discourse on the Stigmatization theme was negative in the early period, but positive discussions were triggered from April 2020 to June 2021. However, negative sentiments dominated the discussion from June 2021 to mid-October 2021. Also, the number of tweets with negative sentiment peaked in July 2021. From November 2021 to the end of the period, the highest sentiment was positive.

In the case of the Role of Government and COVID-19 themes, negative sentiment was mostly expressed compared to positive and neutral sentiments in the earlier period, while positive sentiment dominated the others at the end of the period. The number of negative tweets on the Role of Government theme grew strikingly in June 2020. In the COVID-19 theme, we identified a close gap between negative and positive sentiment from June 2021 to March 2022.

Negative sentiment remained prominent throughout the entire period in Women and Children, Related Diseases, Misinformation, and Outbreak of HIV themes. In contrast, the discussions on Research and HIV Treatment themes were positive for the entire study period.

Correspondingly, in the early period of the study, positive sentiment was prevalent in the HIV Criminalization and Disclosure themes. However, negative sentiment dominated the discussions among users from September 2020 to

February 2021. The number of negative tweets increased dramatically in October 2020 and January 2021. From July 2021 to mid-September 2021, an identical number of tweets was noticed for both positive and negative sentiments. Positive discussions about HIV Criminalization and Disclosure theme became prominent from November 2021 to the end of the period.

Overall, there was a downward trend in the number of tweets for all the sentiments expressed in each theme during July 2020, December 2020, and September 2021. It is observed that neutral sentiment was the lowest sentiment for all the themes over the period.

V. DISCUSSION OF RESULTS AND RECOMMENDATIONS
A. SEXUALITY AND LIFESTYLE CHOICES
On the theme of sexuality and lifestyle choices, the concerns were overly negative toward bisexual men. On this theme, 40.2% of tweets expressed negative sentiments, 32.3% positive, and 27.5% for neutral sentiments. Largely, several concerns were raised by bisexual men being marginalized when it comes to blood donation due to high risk. For instance, a user wrote... **T75:** *It's so gross to me that even in 2020, bisexual men can't donate blood because of HIV prevention when straight people are just as susceptible to contracting the disease. [April 24, 2020].*

Currently, in most countries, there is a deferral period of 3 months since the last sexual contact for bisexual men to be eligible for blood donation. Some men expressed disappointment in being turned away from donating blood because they had sex within the last year, and this made them doubt the FDA's (U.S. Food and Drug Administration) standards for HIV testing.

Also, it was mentioned that when it comes to HIV-related advertisements, people of color within the bisexual men community are regularly the faces shown on TV, which is considered racial bias. It is therefore welcoming to see white men being used for advertisement on a few occasions. This is captured as follows... **T670:** *Thank you Biktarvy for posting a commercial of a straight white male with HIV. [April 4, 2021].*

Furthermore, it has been mentioned that advertisements and commercials on HIV/AIDS have been on the increase lately, but the focus and scope are too narrow on bisexual men. This is considered prejudicial and can lead to neglect. Two users (T10 and T79) mentioned... **T10:** *It's sad that bisexual men people are only represented on TV ads when it's a HIV prep ad. It shouldn't be the only way we see bisexual men couples on TV commercials. [April 29, 2020]. T79: Why is it that when bisexual men couples are featured in a TV series, being HIV positive is always a pivotal point as if straights are immune to it. I get awareness but boxing its effects unto a minority may cause unnecessary neglect from the majority [April 24,2020].*

Recommendation: We recommend that commercials and awareness creation should be targeted to the larger population. A reason for targeting the bisexual men community

could be that the HIV infection rate is higher among them. However, overly targeting them in HIV ads is becoming repulsive, and the good intention could be lost/misunderstood. Besides, it is imperative to focus on the larger population to reduce complacency.

B. TESTING

This theme has about 48.4% positive tweets, 26.6% negative tweets, and 25.0% neutral tweets. Testing remains the most effective tool for early detection and control of the spread of HIV. It has therefore been the focus of the major stakeholders to make testing accessible. According to some, the ease of access to test kits is commendable. This is captured by T5034 and T6166 as follows... *T5034: Took my HIV home test kit this morning which came back positive. It was so easy and now I know my status [February 2, 2021]. T6166: I had an HIV test at a pop-up health fair in a parking lot. It was a finger stick and results took five minutes. No joke. That's how fast and easy it would be to qualify as donor. [February 9, 2022].*

The fact that HIV test kits can be delivered to homes for self-screening has also been dimed as a plus. This is considered by some as a good way to minimize stigmatization and prevent the prying eyes of nurses that are sometimes discouraging. This user says... *T4989: The peace of doing an HIV test at home with no nurse shouting at you or asking why you did not use protection. [February 3, 2021].*

Additionally, the quick turnaround time for test results has improved as several tweets mentioned five minutes window... *T2309: I just got tested for HIV and results were ready in five minutes!!!!!! [August 5, 2020].* Still, turnaround time in other jurisdictions could be longer.

Despite the gains, a major downside/limitation to testing has been the cost of doing so in other jurisdictions. In places with economic hardships, people will not spend on testing to know their HIV status if there is no reason to do so. The following tweets discuss cost... *T482: This company charged me $1780 for a HIV test. Wonder how much they are making off corona virus testing. While they take days and days for the turnaround time. [April 3, 2020]. T1710: I can't believe how expensive the HIV RNA test is. It should be free [April 5, 2022]. T3471: $300 HIV test, it better come back positive for that amount [December 29, 2021]. T15848: A full panel HIV/STD screening is $686; change, America must do better. [September 29, 2021].*

Furthermore, testing is difficult for some people to access. For instance, pregnant women in Nigeria have inadequate access to test kits as captured in this tweet ... *T763: Women coming for the 6 weeks clinic cannot get tested for HIV at this facility because apparently there's shortage of test kits. This country is real messed up [April 20, 2021].*

Recommendation: As a recommendation, there is a need to balance the tides between economies when it comes to facilitating HIV testing. The economies that have test kits in abundance should be willing to donate/share with deprived regions. At the moment, these deprived regions are more than the well-to-do economies so the fight against HIV/AIDS

cannot be worn easily with no access to test kits for the masses. Also, we need to normalize going for HIV tests with our partners regularly or before getting into a relationship.

C. PREVENTION AND MEDICATION COST

The tweets on HIV/AIDS prevention mainly focused on the awareness and benefits of PrEP (Pre-Exposure Prophylaxis). Approximately 48.6% of the related tweets are positive, 35.1% are negative tweets, and 16.3% are classified as neutral tweets. A user was very appreciative of being introduced to PrEP by his partner... *T1293: I got into prep because of her. she was a good doctor. she did my HIV tests. got me into prep. she will always remind me to take good care of myself. her legacy will always live [April 10, 2021].*

The goal of PrEP as a measure to control risky sexual behaviors has been captured properly in most of the tweets. The fact that it is a medication to reduce the risks of contracting HIV has been widely shared. When it comes to HIV and AIDS, PrEP could be used to prevent transmission to the next generation. The following representative tweets captured the following sentiments... *T12837: If you have a risky sexual behavior and are at high risk of being exposed to HIV, kindly get PrEP (Pre-Exposure Prophylaxis), this is a medication that will help reduce the chances of you contracting the virus. [March 21, 2022]. T17435: Instead of being reckless, there is something called PRE-EXPOSURE PROPHYLAXIS AGAINST HIV, which is very effective with good adherence. For all of you who cannot negotiate for condom use, with partners of unknown hiv status, or with multiple sexual partners [September 29, 2021].*

Eventually, it appears the campaign to create awareness around HIV prevention is catching on in some communities. There is a need to encourage people to take and adhere to their medication regularly in addition to the use of other protections such as condoms. Its effectiveness can be shared among experienced users as stated... *T12915: I am planning to go on PrEP, it's a Pre-Exposure Prophylaxis Medicine that people take to prevent getting HIV from sex and adds an extra layer of protection aside from the use of condoms. They say, when taken as prescribed, PrEP is highly effective to prevent HIV. So yeah! [March 19, 2022].*

However, there is room for more campaigns on PrEP as expressed by some. Better still, there should be a renewed effort to evangelize to the masses on PrEP. Some people still have not heard of the medication and its benefits either due to ignorance or simply because they have never been informed. In jurisdictions where sex education has dwindled, people became reckless with their sexual lifestyles and the direct result is an increase in the HIV infection rate. Some concerned users posted these tweets... *T8790: So, few people know about PrEP. I also hate how the aggressive (which was good) HIV/AIDS campaigning has died down, and how I no longer see as many billboards, pamphlets and adverts [March 17, 2022]. T17404: SEX EDUCATIONISTS among the youths are becoming more reluctant on the concept of HIV and AIDS. There have been a continues increase in*

unprotected sexual activities thus there is a rise in infections rate on a yearly basis. [September 1, 2020].

Overwhelmingly, the aversion to this theme is the cost of purchasing PrEP medications. There have also been concerns about insurance companies not covering PrEP medications. The following are some resentments expressed on the cost of PrEP medication... *T8252: Absolutely horrified to find out that the prescription for PrEP after I was assaulted is going to cost me $50 for a three-day supply. And then I'm still going to have to fill the full 28-day supply later this week when my PCP prescribes it. All so I don't get HIV. [March 22, 2021]. T8467: My health insurance company is refusing to cover prep which can reduce the risk of HIV transmission. I live in NYC. What do I do? [March 26, 2021]. T8658: PrEP reduces the risk of HIV by 99%. But with its astronomical out-of-pocket costs, this preventative treatment is inaccessible to the people that need it most. Which is why insurers are REQUIRED to stop charge out-of-pocket fees. [March 24, 2022]. T9044: She was treating a man who had been stabbed, she got cut herself, and ended up with his blood all over her. She needed to take PREP because of the risk of possible HIV exposure. Then she found out her insurance won't cover PREP. [May 29, 2020]. T9081: Imagine being told that the medication you need to take to prevent a possible HIV infection because of being assaulted would cost me $3,500. [May 26, 2020].*

Recommendation: Going forward, it is suggested that PrEP education is intensified as a measure to prevent HIV transmission even among positive partners. Education can start from schools, and it must be made clear to people that protection against HIV infection is a choice. Additionally, sex workers must be sensitized on new preventative measures for HIV/AIDS such as PrEP and PeP. Also, it should be made clear to people that condoms should be used to prevent HIV when PrEP is started because PrEP takes time to achieve its maximal protection - 20 days in blood, 20 days in vaginal tissue, and 7 days in rectal tissue.

Moreover, it will be beneficial if cost-cutting measures are implemented to enable easy access to PrEP medications. As well, insurance coverage should be expanded to cover the cost. And payment exemptions can be made for victims of sexual assault and abuse. These measures can further embolden victims of such circumstances to come forward and open up about their trauma to the appropriate authorities.

D. HEALTHCARE AND SUPPORTING RESOURCES

Under this theme, about 57.4% of tweets were classified as positive, 27.6% as negative tweets, and 15.0% as neutral tweets.

Some agencies are making strides in supporting persons living with HIV/AIDS (PLWHA). This support is also extended to migrant children in some jurisdictions in the form of accommodation and treatment. For instance, an agency in New York, USA tweeted... *T4366: We are committed to providing shelter, care, treatment access and support for people living with and children affected by #HIV/#AIDS.*

This means thousands of #children of mainly migrant workers abandoned by their partners and employers. [December 24, 2021].

The US government has also released funds recently in 2021 to tackle the HIV menace. This type of support is needed especially within deprived communities for community engagement, training, and the purchasing of supplies. A representative tweet reads... *T18234: Last week, the Biden-Harris administration announced more than $2 billion in funding in the fight to end HIV. This includes providing funding to community-based health organizations that serve Latinx communities. A community too often underserved and underrepresented. [October 15, 2021]*

Some private agencies and programs have been designed to support PLWHA. These programs have been hailed for their effectiveness and impact on beneficiaries. Some users tweeted... *T9438: I need a new HIV specialist. I am so thankful to the Ryan White Program for this opportunity to get some needed help. [July26, 2021]. T19223: I live better than most sober people. I am on HASA (HIV/AIDS Services Administration). yes, addict and HIV positive who chooses to live. Thank God I have these programs to help me out [September 12, 2020].*

However, it has been expressed that healthcare and supporting resources for HIV management have not been distributed equitably. As a result, there is a shortage of support. A PLWHA has this to say... *T5297: This year I feel so desperate about HIV support services that I am grateful for this medium to speak out. [December 1, 2021]*

There are also concerns raised about the neglect of rural folks when it comes to HIV support. Mostly, there is limited to no access to healthcare services in these rural areas and in some instances, they must compete with urban areas for logistics and health personnel. This imbalance is expressed globally in both developed and developing economies. Some sentiments read... *T2048: Rural communities continue to be left behind. We must invest in HIV care beyond Atlanta, Houston, and Washington, DC. When I worked in Southwest Georgia, I would inform people about #PrEP but they had nowhere to go to get it. [April 20, 2022]. T19294: I'ma person living with HIV. I work with an elderly person, and we've been struggling since February to get support; someone please help us [September 2, 2020].*

Some people furthered that their government has gotten its priorities wrong when it comes to allocating the budget toward HIV management and prevention. Sentiment on this is shared as follows... *T9867: We do not prioritize #HIV prevention or treatment. Check out the federal, state, and local budgets and you will see the small budgets that are declining year over year. [June 16, 2020].*

Also, the inequities in giving care to persons living with HIV/AIDS have been dimmed as being fueled by racism and marginalization. This is because some people of color feel that compared to their white counterparts, there is little support available for HIV management. A concerned user tweeted... *T3912: The intersection of racism, homophobia,*

ableism, and classism did you say? DING! DING! While HIV preventative measures, like PreP, have gotten better, use and access of those medicines among Black and Latino MSM present as barriers due to discrimination within healthcare and society. [August 2, 2021].

The COVID-19 pandemic has also drastically created a syndemic for PLWHA. As a result, there is competition for healthcare and supporting resources. It was observed that healthcare workers became frontline staff during the COVID-19 pandemic and that has led to limited medical services for HIV treatment. The quality of care for PLWHA has also been affected. This observation is universal. These tweets highlighted the effect of COVID-19 on healthcare support... *T5266: COVID makes it difficult and dangerous for frontline health workers to deliver continuous, high quality HIV services to those who needs them. [December 1, 2021]. T18202: Even after easing of the second lock down, many people living with #HIV especially adolescent girls and young women are still unable to access care services at health facilities. This is a major challenge. [October 16, 2021].*

Recommendation: The agencies in charge of the fight against HIV (e.g., CDHC and NAP) should invest in training a group of influencers on HIV education and resources with a focus on transmission, prevention, and medication. Then, the HIV prevention model should be designed equitably. There should be a clear plan for extending access to people with lower health literacy and reduced capacity to navigate the healthcare system, to scale up without amplifying inequalities.

E. STIGMATIZATION AND VICTIMIZATION

This theme has dominated most literature, so it is not surprising that it featured prominently among the sentiments shared by people on Twitter. The posts on stigmatization, victimization, and discrimination were classified by the machine learning algorithm with the following polarities: 43.3% positive, 41.2% negative, and 15.6% neutral.

It is encouraging to see that some PLWHA are beginning to express confidence in themselves despite their diagnosis and negative public opinions. This show of resilience, bravery, and hopefulness is critical for the fight against the disease and its associated stigmatization. It can also serve as a platform to push for more public education. In this regard, some representative messages read... *T7458: Hey Twitterverse, friends, acquaintances, I am HIV+, undetectable, medicated and have been for 10+ years. I′m making this post for a bunch of reasons; to educate, to combat negative stigma, and to make my status public knowledge. [March 24, 2022]. T2782: I do have HIV, I, am not ashamed at all. I just hope my TRUE story helps somebody else in the long run. [February 22, 2021]. T3221: My HIV status is nothing I′m ashamed of, but comments on here over the last week have shown me that a lot of people still see it as "disgusting" and something to be ashamed of. Undetectable = untransmittable. I have a whole life to live. [February 18, 2022].*

Based on the bravery of some PLWHA coming out publicly about their status, some well-meaning citizens are also showing support. A user posted this message... *T628: People with HIV and AIDS are nothing to be afraid of. They are people just like us, and each has a story to tell. These people should be helped, embraced, and not dismissed. We need to open our hearts and our minds to them, and we just may learn we're pretty much all the same. [April 13, 2021]*

Also, those who are getting proper medical and community support can overcome the stigma; and therapy has also been helpful. A PLWA says... *T4452: Honestly, therapy has changed my whole outlook on life these past couple of months. Even when dealing with being HIV, I no longer feel ashamed about my status. I am still beautiful no matter what! [July 28, 2020].*

It is however important to state that stigmatization, victimization, and discrimination against PLWHA are still endemic. There are uninformed citizens and netizens who still attack and look down upon those who tested positive for HIV. Some people who tested positive are also not aware that there is a treatment for this disease. This level of ignorance causes them to coil and hide. This tweet says... *T1177: I've known too many folks so ashamed of having HIV and just don't know there's treatment. That's how ignorant some folks are today. Lots of STDS are treatable but others want to shame you [August 27, 2020].*

Largely, the bad actions and attitudes towards PLWHA are precipitated by ignorance, lack of understanding, and misinformation. Some messages read... *T1747: The stigma for HIV is bad a lot of you all are very uneducated and I′m tired of it! [August 8, 2021]. T2342: break the HIV+ stigma. it breaks my heart whenever I talk about being born HIV+ and I get people messaging me thanking me for speaking out about it because they're ashamed or scared. It shouldn't be this way. [December 25, 2021]. T3882: I have so many people in my life that I love that are thriving with HIV. I think of them every time I read some ignorant posts. Many of them are too ashamed to defend themselves because of misinformation [January 16, 2021].*

Ignorantly, some think that HIV is limited to lifestyle choices such as homosexuality. Others also erroneously think that HIV is contracted only through reckless sexual activity and/or promiscuousness. Meanwhile, so many people contracted the disease through varied means unfairly and unknowingly. The following representative tweet says... *T4866: We need to break the stigma around HIV/AIDS, people are already so misinformed about the disease. It is not a Gay man's disease; straight people can contract it as well and we need to stop looking at people who are positive as dirty. [July 28, 2021].*

Moreover, stigmatization and discrimination in society can hinder people from testing for HIV. Also, some people are avoiding treatment due to the same reasons. In some instances, HIV-positive status is criminalized, which is pure ignorance and retrogressive. Some users have this to say... *T5851: The problem isn't disclosing your HIV status, the*

problem is how our society deals with it. Many people don't even get tested because they are afraid of any positive result and the discrimination that might follow [June 15, 2021]. **T9700:** *First, even today with the knowledge and U = U people living with HIV still face stigma. HIV is still criminalized. And people are still dying and not adhering or accessing treatment due to this. [September 13, 2020].*

Recommendation: Fighting against stigmatization, discrimination, and victimization of PLWHA will require massive public health education. Most of the tweets posted (whether positive, negative, or neutral sentiments) all blamed ignorance, misinformation, and lack of education as the basis for stigmatization against PLWHA. Through education, people should be made aware that treatment is available, and you can live a normal life if you are HIV positive. You shouldn't be ashamed of your status, and it doesn't make you any less human.

There should be intensified activism and celebrities should be using their platform to educate people instead of being homophobic and spreading false information. Moreover, PLWHA could be encouraged to be brave enough to speak up.

F. VULNERABLE GROUPS - WOMEN AND CHILDREN

Another theme that the classifier identified is the impact of the syndemic (i.e., HIV/AIDS and COVID-19) on vulnerable groups such as women and children in society. The distribution of the polarities in the tweets is as follows - 38.8% positive, 45.9% negative, and 15.3% neutral. These vulnerable groups are at the receiving end of the brunt of all the other themes.

On the brighter side, there is an indication that the mother-to-child HIV transmission rate is decreasing due to regular testing and medications. Mother-to-child transmission of HIV is also called perinatal transmission of HIV. Pregnant women with HIV can take HIV medicines during pregnancy and childbirth to prevent mother-to-child transmission of HIV. A user writes... **T193:** *Due to HIV testing and medications, the number of children infected with HIV during pregnancy, childbirth, and breastfeeding is decelerating. [April 24, 2020].*

Additionally, babies born under high-risk HIV conditions could be given medications to prevent infection. However, this is dependent on the mother knowing her status during the pregnancy. This tweet states... **T1297:** *Babies born to HIV pregnant women are given medications (within 72hrs of birth) to prevent HIV infection...these are taken from 6 to 12 weeks depending on when the women were diagnosed. [April 17, 2021].*

Sadly, however, women and children are highly disadvantaged in the struggle against HIV. There are millions of children with HIV infection with inadequate access to care. Also, most pregnant women are deprived of Antiretroviral drugs (ARVs) due to a shortage of supply or high cost. This problem appears to be global regardless of the economic disparities. A post read... **T1640:** *We are not being loud enough with the ARVs issue. Imagine a baby is going to be*

born HIV positive because the mother could not access ARVs to suppress her viral load and not transmit HIV to the baby. [April 3, 2021]. **T5843:** *Pregnant women with HIV need to have access to antiretroviral treatments to be able to prevent mother-to-child transmission of HIV. [April 14, 2021].*

Additionally, it is becoming increasingly difficult for women to have access to basic preventive measures such as rings. A user writes... **T5766:** *We cannot have access to the ring for HIV prevention because the FDA says, "epidemiology in the US doesn't warrant the ring's approval". but are we dying? Are WE – Black Women- burdened inflicted displaced battling facing the HIV epidemic here in the U-S-of A... yes we are [January 25, 2022].*

Another major issue is the neglect that children, especially girls, suffer from their guardians in society. Some of these girls suffer maltreatment at the hands of people they trusted to care for them. They are either kidnapped, raped, abandoned or unfairly treated in several instances. There have also been situations where these girls are forced into early marriages thereby destroying their childhood. Also, some experts opined that they cannot make a diagnosis of HIV in children, so care is significantly compromised. And without treatment, up to half of the infants infected with HIV die within their first two years. Some concerned users note... **T2588:** *Girl with genital warts and HIV at the same driver's house and she was only 12yrs old. The man was arrested and charged with kidnapping and rape. My heart is absolutely shattered because why would you do that to a child?[April 4, 2022]* **T15299:** *A 19-year-old house girl, forced to marry so called DJ who beats her as often as he wills. At 19 she's got 2 children already and just got diagnosed with HIV [May 27, 2021].*

Furthermore, women and children living with HIV are adversely affected by stigmatization and discrimination. They are naturally disrespected, neglected, tagged as "witches", and victimized. A message reads... **T10969:** *Many of the women who contracted HIV were shunned or disowned by their family and friends. Left to die a horrible death alone, disrespected and abandoned by those closest to them. Never forget the atrocities that were committed at home, for they are the most unforgivable. [June 7, 2021].*

Due to the above reasons, children become caregivers of their parents who are HIV-positive. This concern is exacerbated by the fact that there are limited care providers, misinformation, and miscommunication around HIV. Someone tweets... **T14539:** *Around 20% of orphans and vulnerable children do not attend school regularly and around 18% are sexually abused. HIV also has an indirect impact on children, whereby often they become the caregivers for parents who are living with HIV. [May 22,2020].*

Recommendation: Stakeholders need to allocate more resources to assisting vulnerable groups (i.e., women and children) in the fight against HIV/AIDS. There are reports of babies and children in orphanages with HIV, and they need assistance. Moreover, pregnant women need to be tested for HIV and they should be encouraged to take their antenatal

classes, as they are paramount to the pregnancy. Also, children who are born exposed to HIV must have a point of care early infant diagnosis. Schools should be supported to enable children to learn about the risks and management of HIV. Finally, public health education should emphasize the de-stigmatization of the disease, especially for the vulnerable.

G. ROLE OF GOVERNMENT

The tweets also reveal several mentions and direct messages to and about governments with their responses to the HIV/AIDS menace. The machine learning classifier considers the related tweets to be grouped into the following polarities - 33.4% positive, 53.4% negative, and 13.2% neutral.

Though the sentiments towards governments on their role in the management of HIV are overly negative, there are a few bright spots. There is the mention of under-developed economies receiving HIV drugs as aid from advanced countries. A post by a user says... *T125: HIV positives can live a normal life now, on free drugs paid for by foreign governments and agencies. [April 10, 2020]*

There is also progress being made by some countries to purchase more antiretroviral medications for their citizens and further stockpile their depleting inventories. This is encouraging because some of these countries have experienced shortages of supply. A representative post says... *T246: People living with HIV can breathe a sigh of relief after the government acquired ARVs following months of shortage, MOH says. [April 8, 2021]*

Moreover, efforts are being made to get HIV drugs to PLWHA at reduced prices. Considering the current economic hardships and high inflation prices, this gesture is deemed welcoming. A user posted... *T989: HIV medication is so expensive and yet the government has already subsidized it. What is done is done. But please use the meds wisely, the money for those meds can be used for other diseases and research. [February 14, 2022]*

However, there is a public outcry against some governments regarding their management, storage, and distribution of drugs. For instance, Kenyans are outraged about the government distributing expired HIV medications to the citizens. Some users said... *T182: Waking up to news that the ministry of health distributed toxic/expired ARVs to HIV/AIDS patients. Uhuruto government is killing Kenyans. [April 28, 2021]. T183: How many people living with HIV/AIDS are in this country? Does the Ministry of Health and the government as a whole care about them? Why give 'poisonous' (toxic) ARVs to the public? [April 28, 2021]*

There are concerns also about misplaced priorities and PLWHA being ignored by their governments. The people of Ontario, Canada expressed their misgivings about the government ignoring PLWHA under the pretext that they are not at risk. Due to this miscalculation, HIV-positive patients were at high risk of the new COVID variants. A user says... *T226: The excuses for Ontario completely ignoring HIV+ are pathetic. "HIV+ are not at higher risk". They don't know that. That's a guess based on inconsistent data. HIV+ are*

vaccinated in other places. We have no idea how HIV+ will manage with the new variants in Ontario. [April 12, 2021]

Also, the scarcity of HIV drugs has been blamed on high taxation and corruption by governments. Some importers are not able to pay the port charges required to clear the drugs from the harbor after they arrived. Some users write... *T269: HIV patients are in distress following acute shortage of ARVs. The drugs are stuck at the port as government has slapped sh90m tax on the consignment that's supposed to be distributed for free by donors because they bypassed corruption-ridden KEMSA [April 3, 2021]. T270: Now the government is creating an artificial shortage of the life-saving ARVs medication by withholding the crucial drugs until donors pay tax. [April 3, 2021]*

There was also a negative impact on the distribution of HIV medications during the electioneering period in the USA for instance. Medication shipments were delayed due to the high volume of mail ballots. Also, some politicians have been blamed for using bisexual men as the reason for the spread of HIV. This misinformation has led to hatred for these men. Some representative tweets read... *T437: To everyone I was just told by my new pharmacy they are having issues getting people their HIV medications. The postmaster general says the delay is due to mailing voting ballots [August 17, 2020]. T756: The government is the reason the HIV and AIDS is attached to Gay men. They put this misinformation out there so everybody can attack one another. [December 10, 2021]*

Recommendation: Governments need to get their priorities right on public health issues. They could look into hiring temporal workers during high demands on the postal services. Also, issues with high taxation should be considered by governments to make the drug affordable and available. This can prevent artificial shortages from being reported in certain regions. Also, appropriate agencies should be funded to facilitate a steady supply of medical resources needed in the fight against HIV/AIDS. There must also be a deliberate effort by governments to fight corrupt officials who squander monies meant to fight the disease.

H. RELATED SICKNESS AND DISEASES

The machine learning classifier also featured diseases and sicknesses that PLWHA are highly susceptible to as a theme for consideration. Approximately 30.0% of the tweets were grouped as positive, 52.0% as negative tweets, and 18.0% as neutral tweets. This theme underscores the fact that HIV infection comes with other medical conditions. Messages posted expressed the fact that many people with HIV still develop Opportunistic Infections (OIs) because they may not know they have HIV, they may not be on treatment, or their treatment may not be keeping their HIV levels low enough for their immune system to fight off infections.

This theme is mostly skewed toward negative sentiments. The few positive tweets mainly highlight the fact that being on HIV medications has helped boost their immune systems against COVID-19. Further, it has been noted that with increasing access to effective HIV prevention, diagnosis,

treatment, and care, it has become a manageable chronic health condition, enabling people living with HIV to live long and healthy lives. A representative post read… *T18715: I am a long-term HIV+ 61-year-old male who has either not got #coronavirus or had a mild case that presented as a normal cold. To my knowledge, whoever is in on #HIV meds has good immune system. [March 27, 2020]*

There have been mentions of reactions (i.e., side effects) of some of the HIV medications. A PLWHA wrote… *T2094: The drug called Biktarvy is an experimental nightmare that should be illegal. I have experienced every single side-effect down to the horrid nightmares and awful sweats. I was told that it's normal to have night sweats due HIV. [April 24, 2021]*

Moreover, there are some PLWHA who experienced sicknesses such as bronchitis, cold, flu, pneumonia, TB, cancer, and diabetes. A user writes… *T3466: Had COVID-19 2yrs ago, had cold within 3 days and it was bronchitis, and I was hospitalized because I was having a hard time breathing even with an undetectable viral load. Having HIV for so long has affected my lungs. [March 8, 2020]*

Most of the tweets however mentioned the toll on mental health for PLWHA. A lot of them experienced depression, anxiety, high blood pressure, stroke, low self-esteem, and so on. Some representative tweets read… *T5116: People living with HIV often suffer from depression and are burdened by several (lifesaving and incredible) pharmacological treatments to manage their disease. [September 29, 2020]. T8581: Although diligent use of medications could prevent the progression of HIV to AIDS, the disease cannot be cured. Victims of AIDS often deal with stigmatization, shame, low self-esteem, and mental health illnesses. [December 3, 2021]. T5707: I am HIV+ and living with the virus for 3 years now. The first half of 2021 was tough for me. Experienced a lot that made me feel so lost. Heartaches, betrayal, self-harm, mental health issues you name it. [June 27, 2021].*

Recommendation: It is our position that caregivers are familiar with these symptoms and associated diseases that affect PLWHA. However, a lot more could be done about explaining the side effects of the prescribed medications. Also, persons who tested positive for HIV could be informed about the possible symptoms and sickness. More programs/activities could be organized such as therapy sessions to alleviate depression and anxiety.

I. HIV CRIMINALIZATION AND DISCLOSURE

This theme essentially focused on the sentiments surrounding criminalizing HIV, consent, and disclosure. There were approximately 45.1% positive tweets, 40.2% negative tweets, and 14.7% neutral tweets.

A lot of people find it useful, humanly, and morally correct to disclose their HIV status with partners. Preferably, at the beginning of the relationship even if you are at the undetectable level. The disclosure could be done when the person living with HIV assesses his or her situation and believe it is safe to do so. However, it is strongly encouraged that disclosure precedes any sexual or risky activity. A user wrote … *T6: Disclosing your HIV status to your sexual partners is important. I personally do it in the beginning of the relationship, some prefer to do it after some time of getting to know a person. [April 28, 2020]*

This view is further strengthened by informed consent. Some PLWHA fear stigmatization and discrimination and as a result, will keep their status private. Several users concede that not every person living with HIV is out to get someone or spread the disease thus, will hide their status from their partners. A concerned user wrote… *T3282: Surely, we can talk about informed consent without dehumanizing and demonizing people who have HIV. [June 15, 2021].*

Moreover, jurisdictional laws give PLWHA the choice to either disclose their status or not if they are at an undetectable level for a certain period. This is because it is perceived that undetectable means untransmittable. Despite the existence of these laws, it is still recommended that informed consent morals are upheld. A post reads… *T1054: In the state of NC, a person living with HIV does not have to disclose their status if they maintain an undetectable status for six months or more. I still let my sexual partners know because it's better to educate them than to let them live with the false stigma surrounding HIV. [August 12, 2021].*

It is important however to note that in the USA for instance, in 21 states, laws require people with HIV who are aware of their status to disclose their status to sex partners, and 12 states require disclosure to needle-sharing partners. Some users also opined that PLWHA knowing they are undetectable makes it easier to share their HIV status with others, as it can be reassuring for others to know that your health is protected, and you can't pass it on too. There is also a legislative push to introduce laws that will not require PLWHA to disclose their status if they are undetectable. It is believed that such legislation will further reduce stigmatization and discrimination. A tweet read… *T4457: The PM has introduced the Infectious Diseases Act (HIV Amendments) Act to Parliament. If passed, this landmark bill will delete all federal requirements on STI disclosure and reporting in sexual intercourse for persons with undetectable HIV, in hopes of reducing stigma. [May 30, 2021]*

The responsibility of knowing has also been placed at the doorsteps of everyone and not just the PLWHA. Some posited that everyone must make the right choices from an informed position. According to a user… *T4550: I have been fighting with people for over twenty years regarding the right to privacy versus the public good. I am HIV+ and my right to keep my health status private should be as important as your right to make informed choices. [May 19, 2021]*

However, criminalizing HIV has received widespread condemnation. Initially, criminalization was meant to curtail the spread of the virus but, it has rather created a culture of silence for the PLWHA. Some said criminalizing HIV infection is simply a way of discouraging people from knowing their status. A user wrote… *T1755: The original intent of HIV Criminalization laws was to reduce transmission of HIV, but*

these laws backfired by incentivizing ignorance of one's HIV status (avoiding testing/diagnosis), which creates a significant public health issue. [February 24, 2021]

There are several hashtags on the fact that HIV is not a crime and should not be perceived as such. E.g., *#HIVIsNotACrime*. Rather, these criminalization laws are perceived to be discriminatory and rather a recipe for fueling stigmatization and public discrimination. Particularly, compelling people with undetectable viral loads to disclose their status has been deemed to be victimizing and offers no tangible protection for partners. A user wrote... *T721: The criminalization of HIV transmission will increase stigma and discrimination against persons living with the virus. It will drive PLWHA underground; persons will be reluctant to get tested for HIV and violence against PLWHA will undoubtedly increase. [June 15, 2021]*

Nevertheless, attempts to relax the current rules have also been met with resistance. For instance, advocacy to decriminalize intentional infection has been described as an anti-social norm. A user said... *T5709: By the way, Senator Wiener advocated in 2017 to remove the felony penalty for knowingly exposing another person to HIV. Come on California, stop pushing this perverted, social progressive agenda that goes against any reasonable social norms. [September 16, 2020]*

Recommendation: There is a need to encourage people to disclose their status but in a safe way. Existing laws that have become counter-productive should be amended or abolished to allow for free information consent and disclosure. Also, the public needs to be educated on certain critical aspects of handling issues with PLWHA. For instance, what needs to be done by companies when their new hires tested positive for HIV? Also, the public needs to know what to do when someone is suspected or caught intentionally spreading HIV.

J. HIV/AIDS RESEARCH

A total of 6.0% of the total tweets collected discussed HIV/AIDS-related research. The research theme is classified with the following polarities: 44.1% positive, 34.8% negative, and 21.2% neutral. The issue of research has been suggested to address the multifaceted nature of the HIV menace.

For instance, undertaking research could help understand the imbalance between offering care to adults and children living with HIV. This can further aid in understanding how to better treat vulnerable groups by professing lasting solutions. Research could help understand the ethical costs of excluding pregnant and breastfeeding populations at risk for HIV in PrEP clinical trials and promote inclusion to maximize the benefits of PrEP tools in the pipeline. On care disproportionality, a user wrote... *T185: Barriers to care disproportionally impact children living with #HIV. That is why we work with research access organizations to develop solutions to overcome these obstacles - bringing us closer to our goal of leaving no person with HIV behind. [April 23, 2020]*

Several tweets also deem research as the best bet for the development of effective drugs and vaccines against HIV. HIV surveillance could also advance the current state of treatment and management of the virus. Some of the promising studies and surveillance in this area are focused on mRNA and HIV genetic sequencing. Meanwhile, years of HIV research have established a profound understanding of viral surface proteins and how this affects immunity, resulting in a detailed understanding of how antibodies target these viral proteins. Some users posted... *T1338: Really hope this HIV vaccine is the real deal. mRNA research might be the game changer humanity needs to best the worst. [April 7, 2021]. T5284: It is good and encouraging to see that key organizations in the U.S. HIV movement are keeping up the heat on molecular HIV surveillance as it is currently being rolled out, and that the HIV press is continuing to cover the issue. [January 22, 2021]*

The commitment of governments and donors to fund HIV research has also been featured prominently in the tweets. These funds can be used to attract and retain the best researchers, as well as to purchase the necessary logistics. Some users wrote... *T2029: Under this partnership co supplied TLD for LMIC countries and participate in tender. Major funding is for HIV and malaria; funding has increased this year. Need to track on the funding of these 2 organizations as they work in 3-year cycle [August 23, 2020]. T6698: There is no cure. There is no vaccine. Perhaps it's time to invest more funds into HIV research, prevention, care, and treatment so that we can end the HIV epidemic. [June 16, 2020]*

However, there are ethical and moral concerns raised regarding the development and testing of HIV medications. There have been ill-informed suggestions that drew the anger of many; the suggestion that humans from developing countries be used to test drugs. For instance, this representative tweet summed this repugnant position... *T664: Missing from the outrage to the Racist French scientists' comments on using prostitutes in Africa for research in HIV is the fact that YOUR leaders approve such studies on their people [April 8, 2020].*

There are also calls to fund surveillance in poor economies where HIV treatment is limited. Most of these poor economies have limited labs though they have high prevalence cases. A tweet read... *T11774: A lot of LMICs entered the pandemic with limited lab and ID surveillance systems, except for TB. Even HIV surveillance is quite limited. Hopefully COVID will drive investment that builds capacity that can benefit reporting for other communicable diseases. [October 20, 2020]*

Recommendation: On this theme, the consensus is that research and surveillance are necessary to advance treatment and care for PLWHA. There are therefore calls to fund more HIV-related research works with greater urgency. Also, biotech and pharmaceutical giants need to realize that when it comes to HIV, monopolizing the patent and keeping trade secrets should not be encouraged in the industry.

K. MISINFORMATION

The machine learning model reported topics within the tweets that have been classified as misinformation to the public. The misinformation can be misleading and prejudicial. Approximately 34.6% of the sentiments were positive, 49.4% were negative, and 16.0% were neutral.

On the brighter side, some found the public health messages during the pandemic educative and enlightening. Some people mentioned that they learned the difference between HIV and AIDS with one being the virus and the other being the disease. A representative tweet reads... *T28: One good thing to come out of the last couple of months is that I've learnt a lot about viruses. There is a difference between #HIV virus and #AIDS. [April 29, 2020]*

There was also a huge push to let the world understand and know that the advent of the COVID-19 pandemic should not make us forget about other existing pandemics. Also, stakeholders and political leaders have been called upon to appreciate the fact that equal emphasis is placed on all the pandemics, which has created a syndemic. A user wrote... *T3376: people keep saying "in the middle of a pandemic" but can we start saying "in the middle of two pandemics" because HIV hasn't stopped being pandemic. [August 18, 2020]*

There were also different perspectives and ideas shared on the impact of vaccination in dealing with the syndemic. A lot of the misinformation centered around the side effects of the COVID-19 vaccination on PLWHA. However, some have posted experiential messages to counteract those fears and worries. Someone tweeted... *T5414: people will believe that the vaccine is causing HIV or side effects; I know someone personally that took it and is fine. [December 30, 2020].*

However, there were unbelievably misleading tweets about the vaccines. Some think that the COVID-19 vaccines contain HIV. Others believe that COVID-19 has an HIV factor in it. A tweet reads... *T22: Viruses do not spread like bacteria. Vaccines will risk you to viruses like HIV and Corona Virus. [April 29, 2020]*

There were also several racial and biased tweets too about race and HIV. Some users think that people of a certain race are all PLWHA. This kind of uncouth behavior calls for message governance and the enforcement of decorum on social media platforms. A user wrote... *T2074: I just had a white guy tell me that all black people have HIV. I was so confused as to why someone would just say that for no reason. These white guys need to be monitored on these social media platforms. They enjoy bothering people with weird stuff. Racist ideas annoy me. [April 29, 2022]*

Recommendation: There is a need for a concerted effort on the part of the public and moderators of public forums to control the spread of misinformation. This should however not compromise freedom of expression and free speech. Blatant misinformation and racial innuendos could be flagged or deleted. Currently, the Twitter platform flags posts that are politically inaccurate and hateful but there is a lot of misinformation on HIV that has not been flagged.

Moreover, public and civic education measures need to be intensified to sensitize the public to the myths surrounding HIV.

L. OUTBREAK OF HIV/AIDS

There were tweets on the outbreak of the HIV/AIDS pandemic as well. About 31.1% of the related tweets were positive, 47.9% negative, and 21.0% neutral.

Some users believed that the COVID-19 pandemic aided in the reduction in the spread of HIV. For instance, the lockdown was considered a good initiative that controlled reckless activities. Someone wrote... *T37: While Extended Community Quarantine is mitigating the spread of #COVID-19, it's also curbing the HIV epidemic in the Philippines- less access for both injecting drug users and for those with risky sexual behaviors. [April 28, 2020]*

Some users also suggested that the same programs that were rolled out to control COVID should be used to control HIV. However, that is not advisable because HIV does not spread like COVID. For instance, there were calls to quarantine infected people, lock up those that were in the window period, retest, treat, educate them on how to behave, and give them life-supporting pills. However, these kinds of treatments should not be proffered for PLWHA.

Moreover, the numbers did not show a global balance in infection rate decrement. Some regions experienced high infection rates. Areas such as Maryland and West Virginia in the USA experienced high infection rates that were unexplainable. Similar trends were noted in other parts of the world such as South Africa. In the case of the latter, it was said that people were simply misinformed. Some representative tweets read... *T421: You all really are not taking this Virus serious and wonder why Baltimore STD and HIV rate is high. [April 6, 2020]. T743: HIV and STD rates are high in South Africa because people believe if you're circumcised your chances of getting it are slim [April 4, 2021].*

Others also expressed that the rise in infection rate is a result of stakeholders focusing too much attention on the COVID-19 pandemic at the expense of other pandemics. This sentiment has featured prominently in other themes too as the reason for the neglect of PLWHA. A user wrote... *T6398: People forgot about HIV while focusing on Covid-19 now HIV stats are high. [October 11, 2020].*

Recommendation: As already seen with other themes, HIV/AIDS stakeholders should dedicate equal attention to the fight against HIV.

M. HIV TREATMENT

There were lots of discussions also on the balance of treatment of HIV which has been classified as a theme. Overall, approximately 46.5% of tweets were considered positive, 30.3% were negative, and 23.2% were considered neutral. On the bright side, some PLWHA expressed their satisfaction with medical breakthroughs in the treatment of the disease. There were thousands of tweets confirming the efficacy of

the prescribed medications due to the reduction in viral loads. The main point is undetectable level is equated to "untransmittable", which is a huge win against the spread of HIV infection.

There were some PLWHA who had high viral loads and after taking medication are now in the undetectable phase. Some testimonial tweets read... *T981: Five years ago this week I was diagnosed with HIV with a high viral load and a low CD4 count. Today thanks to my meds I am undetectable and my CD4 levels are great [August 1, 2020]. T1216: After almost a year of antiretroviral treatment, I can finally say not only am I undetectable, but my immune system levels are back to normal/average of someone not infected. I could cry of happiness. [December 30, 2020]. T3564: I am HIV Positive and have been Undetectable for almost 12 years now. To our community - please educate yourselves. Undetectable = Untransmittable. [March 28,2021]. T3730: It's been 14 months since I was diagnosed as being HIV+. I only started treatment 6 months after I was diagnosed and now I am undetectable and feel healthy. I am now over the shock and feel happy. [March 7,2021].*

What is also encouraging about the above tweets is the sentiments of joy and hope expressed by many. Clearly, properly taking medications has been shown to reduce viral loads to undetectable levels. This could well mean that HIV diagnosis is no longer seen as a death sentence as it used to be in the past. The presence of medication has made the treatment journey and acceptance of the disease easier for many PLWHA. It is also good to see people giving varied time frames within which their status moved to undetectable levels. This is necessary to reduce anxiety for someone who has been newly diagnosed.

Some PLWHA shared how quickly they saw encouraging turnaround from the time of diagnosis to reaching an undetectable phase especially when they do not miss their medication. The medication essentially keeps the virus from replicating. There were instances where it took only three months of medication to be undetected. However, this was based on early detection. A user wrote ... *T4818: Before it comes out from anyone else, I feel I should address something that has happened to me. In May, I was diagnosed with HIV. Luckily, they caught it early, and by the beginning of August, I already reached undetectable, where I will remain as I stay on my meds. [October 12, 2020].*

However, the global fight against HIV in terms of medication supply still has a long way to go in terms of accessibility, affordability, and availability. It has been reported that nearly 53% of children living with HIV do not have access to medication. A user wrote... *T316: Still mind blowing that only 53% of children living with HIV have access to drugs and only 38% are virally suppressed. This is a crisis. [April 23,2021].*

In places like Uganda, a significant number of PLWHA whose treatments were disrupted by the COVID-19 pandemic did not continue with treatment. This trend has been seen in other economically strained regions globally. A post says... *T330: Gulu District has failed to enroll back 9,000 people living with HIV on ARVs after abandoning treatment during the nationwide lockdown due to COVID pandemic. Only 18,000 out of the 27,000 people who tested positive for HIV in the District are still on treatment [April 24,2021].*

Also, when the diagnosis is late, it makes treatment longer and more complicated. There were concerns that some minority groups (e.g., South Asians in the UK) may be subjected to late diagnosis. Furthermore, late diagnosis incidents have also been widely reported. This is because most people do not get tested early or get tested at all. Some users posted the following ... *T1787: South Asians are living with HIV in the UK. Like other racially minoritized groups, they are more likely to be diagnosed late with HIV, meaning they could have tested earlier [February 5,2021]. T2110: Late diagnosis rates are still incredibly high. In 2019, 42% of people were diagnosed late with HIV. For some, this can mean poorer long-term health outcomes, leading to an eightfold increased risk of death. Clinical reviews show many could've been tested sooner. (Source: PHE) [January 29,2021].*

Affordability has also been an issue for many. Simply put, some people cannot afford the cost of treatment. Some wrote... *T3045: These days, we continue to see HIV/AIDS doing the most harm in the Global South, where, despite the efficacy of antiretroviral drugs, treatment continues to be too expensive. [June 5,2021].*

In cases where there are distribution and delivery delays, PLWHA can be adversely impacted. For instance, when postal services do not deliver medications on time, there is the possibility that a PLWHA can or may develop viral resistance. This is an issue that can mainly affect rural folks even in advanced economies. Someone wrote ... *T3733: I'm really fed up with the United States Postal Service not delivering my HIV meds on time. 4 days now without it and I'm worried about viral resistance. [March 27,2021]*

Also, there are so many PLWHA who are not on effective treatment. This trend has been reported in several countries. A representative tweet reads... *T4969: Almost two-thirds of transgender people in India have no access to treatment for sexually transmitted diseases. While 7.2% of them are living with HIV, only 59% have been referred for testing, and a mere 33% have received counselling. [October 26,2021]*

Recommendation: Clearly, it has been established that people who regularly take their HIV meds have experienced good results in viral load reduction to an undetectable level. Indeed, many people living with HIV who adhere to regular antiretroviral therapy (ART) maintain undetectable levels of the virus in their blood. This is a hopeful sign knowing that HIV is now manageable with treatment. However, some people are in denial and do not want to face the reality of their diagnosis. It is our general belief that with advancements in medication and good testimonials, doubters and deniers will accept treatment. Moreover, the good outcomes of taking medications regularly will boost confidence in people to embrace the idea of early testing. In places where there are system failures, stakeholders need to introduce interventions to

educate and engage affected communities on how to provide support.

N. COVID-19

Even though the themes considered throughout the study have been influenced by COVID-19 regardless of the sentiment in the post, there were about 30,991 tweets (representing 14.46% of the total data) that are classified under the COVID-19 theme. Within this theme, 40% of the tweets were positive sentiments, 41.6% were considered negative, and 18.4% were neutral.

Generally, people were excited about the early breakthrough in finding vaccines to treat COVID-19. The good news for PLWHA is the fact that similar findings could fast-track the discovery of vaccines for HIV. A user wrote .....
T27118: I'm excited for the potential of mRNA vaccines. This is a new technology that's being deployed here but despite that, the results seem very hopeful. There are also clinical trials that are working on an mRNA vaccine for HIV as well. [November 20, 2020]

Some users have problems with the fact that an international independent body has not been established to investigate HIV even though the same has been done for COVID-19. Compared to the COVID-19 vaccines, PLWHA expressed dismay at the cost of getting HIV vaccines.

VI. CONCLUSION

Several people are living with HIV/AIDS (PLWHA) according to data from the World Health Organization (WHO). Even before the outbreak of the coronavirus disease 2019 (COVID-19) pandemic, the HIV pandemic could not be effectively contained globally. As a result, the combination of the two pandemics has created a *syndemic*. Therefore, the goal of this research is to study and highlight the issues affecting PLWHA during this period of global medical disruption and emergency. We approached this objective by applying machine learning techniques to collect textual data from social media.

To achieve our goal, we focused on the Twitter social media platform and extracted approximately 2.80 million tweets related to HIV/AIDS posted from March 01, 2020 to April 30, 2022. The total dataset was cleaned, and duplicate tweets were removed to obtain a unique dataset of over 200 thousand tweets that were used in this study. Then, sentiment analysis was done to determine the following three polarities in the tweets: positive, negative, and neutral. Additionally, stopwords and punctuations were maintained during the sentiment analysis since they contribute to the overall polarity of texts. However, for the topic modeling, we converted the tweets into lowercase and removed punctuations and stopwords. Finally, the tweets were lemmatized using the spaCy library and converted into a document word matrix to discover the topics.

We discovered 25 HIV/AIDS topics using the Latent Dirichlet Allocation (LDA) topic modeling algorithm, and then the related topics are categorized into 14 *themes*. Then, thematic analysis was done to identify the positive and negative factors influencing each theme using VADER and TextBlob sentiment analysis tools.

A. MAIN FINDINGS AND RECOMMENDATIONS

Based on the thematic analysis performed with the aid of a machine learning algorithm, we identified fourteen themes that were deemed as issues affecting PLWHA. The distribution of the themes has been summarized visually in Figure 15.

Policymakers need to understand that PLWHAs are facing a syndemic (which is the occurrence of two pandemics at the same time. In this case, HIV and COVID-19. Thus, there is a need to focus on the equitable distribution of health supplies to these groups. Policymakers are urged to engage all relevant stakeholders to fight against stigmatization, criminalization and disclosure, misinformation, and the management of the cost of care.

The overall findings from the work are as follows.

(*a*) *Sexuality and Lifestyle Choices* – On this theme, bisexual men felt marginalized when it comes to blood donation. Also, people of color within the bisexual men community are mostly used for HIV/AIDS television and other media advertisements which leads to damaging prejudice. Moreover, television ads are presented to target the bisexual men community. The net effect of this is racial bias and negative prejudice towards bisexual men. It is therefore recommended that media ads target the larger population to reduce public complacency on the issue of HIV.

(*b*) *Testing* – As testing is considered fundamental for the early detection of HIV, it is highly recommended that test kits are made readily available. In some places, kits are delivered to homes and quick turnaround times for testing results have been reported. However, so many people complained about the high cost of testing. In other places, people do not have adequate access to test kits. Furthermore, society needs to normalize HIV testing which can aid to control the spread of the disease.

(*c*) *Prevention and Medication Cost* – Most of the users accurately captured the benefit of Pre-Exposure Prophylaxis (PrEP) which is a measure to control risky sexual behaviors and transmission. The creation of HIV/AIDS awareness is catching up and people are getting the message. However, some people have not heard about PrEP. Also, the high cost of purchasing PrEP medications has been highlighted by many users with some expressing that the medications are not covered by medical insurance.

(*d*) *Healthcare and Support Resources* – There are groups, agencies, and institutions that are extending support to PLWHA through the financial support of governments and donors. Programs are being developed to assist PLWHA. However, there is a massive shortage of support in so many places globally, especially in poorer economies. As well, rural folks feel neglected with HIV support while some minority groups feel marginalized. These challenges are further complicated by the competition for medical supplies and support created by the syndemic (i.e., the simultaneous outbreak of HIV/AIDS and COVID-19 pandemics).
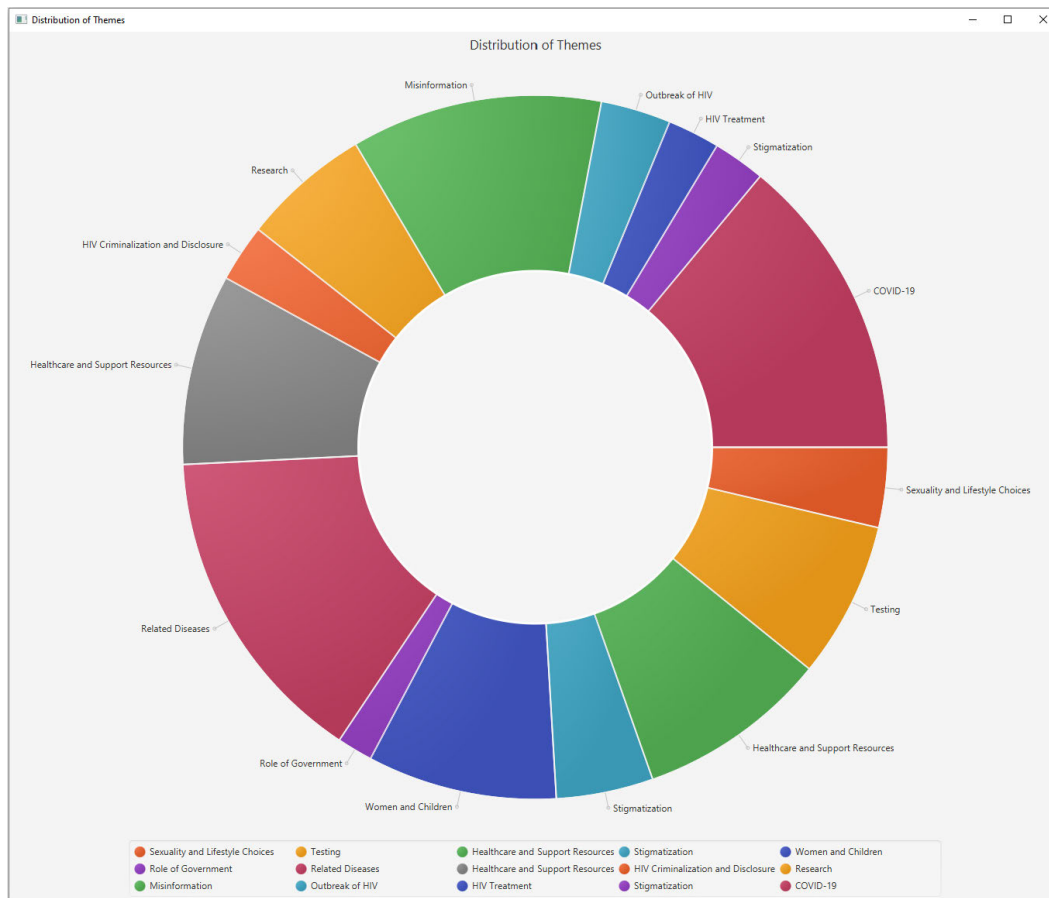
FIGURE 15. Distribution of themes in the tweets.

(e) Stigmatization – Some PLWHA are open about their diagnosis without fear of stigmatization, discrimination, and victimization. As a result, the public in some instances posts messages of support and encouragement in return. However, stigmatization of PLWHA is still very much endemic where people post negative and insulting comments. These bad actions we found could be attributed to ignorance and misinformation. A recommendation to overcome stigmatization is to embark on public education on HIV/AIDS.

(f) Vulnerable Groups (Women and Children) – There are some indications that the mother-to-child HIV transmission rate is on the decline due to regular testing and medication in well-to-do regions. However, the most vulnerable groups are women and children with millions of infected having no access to medical care. Some pregnant women are also deprived of Antiretroviral drugs (ARVs) while some women could not access basic preventive measures such as rings. In situations where diagnosis could not be carried out among children, care is compromised. These vulnerable groups are also the most affected by stigmatization, victimization, and discrimination.

(g) Role of Government – Some governments are doing well while most of them are not in the fight against

HIV/AIDS. Some countries in poorer economies are purchasing more antiretroviral medications for their citizens. However, most users expressed disdain for their government's response to the HIV pandemic. There were expressions of anger towards governments on expired drugs being distributed, priorities being misplaced, scarcity of drugs, and poor distribution mechanisms.

Policymakers need to understand that PLWHA are facing a syndemic (which is the occurrence of two pandemics at the same time. In this case, HIV and COVID-19. Thus, there is a need to focus on the equitable distribution of health supplies to these groups. Policymakers are urged to engage all relevant stakeholders to fight against stigmatization, criminalization and disclosure, misinformation, and the management of the cost of care.

(h) Related Sicknesses and Diseases – Some PLWHA develop Opportunistic Infections (OIs) because they may not know they have HIV while others reported side effects of their HIV medications. A lot of the PLWHA also expressed concern about mental health as they experience depression, anxiety, and low self-esteem. Caregivers who support PLWHA must familiarize themselves with these associated side effects.

FIGURE 16. Word clouds of tweets for 14 themes.

(i) HIV Criminalization and Disclosure – Some users commended laws that require the disclosure of one's HIV status with their partner. However, some states allow PLWHA with undetectable viral loads to disclose their status by choice. However, criminalizing HIV has been widely condemned. That is because some people just end up discouraged from taking HIV tests. Criminalization laws have been discriminatory as well. As a result, existing laws that have become anti-progressive should be amended or abolished.

(j) Research – This theme is hailed as the best way to quickly develop vaccines and treatment options. It could help understand the imbalance in support for PLWHA as well as how to engage the vulnerable in the community. Thus, a lot of users called on governments to fund HIV/AIDS

research. Research however should not compromise ethical expectations for drug testing for instance.

(k) Misinformation – Several tweets are simply misinformation, misleading and prejudicial. While other messages posted on public discussion forums are educative, others were simply negative propaganda. There were misinformed tweets on race, culture, and gender. It is recommended that moderators of these public forums introduce measures to sanitize messages posted or flag them as inappropriate. At the same time, freedom of expression should not be denied.

(l) Outbreak of HIV – In some instances, some measures introduced to control the COVID-19 pandemic (such as lockdowns) aided in the reduction of the spread of HIV. However, some regions experienced an increment in infection rate

during the same period. It is acknowledged that the rise in the HIV infection rate could be attributed to overly focusing on the COVID-19 pandemic.

(*m*) *HIV Treatment* – There were several PLWHA who confirmed the effectiveness of the prescribed medications for treatment as they experienced a reduction in viral loads. The treatment time for experiencing a positive outlook however differs from person to person. Some people said their viral loads reduced within three months while others took longer. It has also been reported that regularly and properly taking the meds is necessary to reach undetectable levels. However, several PLWHA globally is constrained by issues such as inaccessibility, unaffordability, and unavailability of medications.

(*n*) *COVID-19* – The general belief among the PLWHA is that the speed with which COVID-19 vaccines were developed could be emulated for the development of HIV vaccines. Also, some users call for the establishment of an international independent body to investigate and oversee the management of HIV.

B. LIMITATIONS

There are some limitations to our work that are worth mentioning. Firstly, because we considered only English texts for the final analysis (since we eliminated all non-English tweets from the total number of collected tweets), the views considered ultimately are from the regions and economies that primarily communicate in English. This means that the sentiments, views, and concerns expressed by the PLWHA, and other users are reflective of mainly English speakers and writers. From the data, as shown in Figure 3b, the representative tweets are mostly from *the Anglosphere*.

Secondly, despite the high performance of the VADER sentiment analysis tool (with 90% F1-Score), we found that the tool did not do well with *sarcasm*. Some tweets were classified as positive polarity tweets that were negative (and vice versa) because the writers used sarcasm to present their opinion. It is important to note however that this point is minimal and did not adversely affect the final analysis, conclusions, and recommendations.

C. FUTURE WORKS

The immediate step going forward is to develop a website/application that uses visualization to display the main issues affecting PLWHA based on geographical locations. The application will be interactive and engaging. This application will be beneficial to all stakeholders, especially policymakers to see areas that need urgent attention and prioritization in the fight against HIV/AIDS.

Also, we aim to transition the data collection step from batch data to using data streams. In this case, we will be able to generate the visualization mentioned above based on live tweets. This will even heighten the voice of community members in real-time.

Additionally, we will leverage machine-translation tools to translate non-English tweets into English to track the insights of non-English HIV/AIDS tweets. In addition, the period of the tweets collection will be extended to investigate how the themes discovered in this work evolve. Lastly, we will train Deep Learning models for the classification of sentiments.

APPENDIX
WORD CLOUDS FOR 14 THEMES
See Fig. 16.

REFERENCES

[1] A. Caputo, A. Giacchetta, and V. Langher, "AIDS as social construction: Text mining of AIDS-related information in the Italian press," *AIDS Care*, vol. 28, no. 9, pp. 1171–1176, Sep. 2016.

[2] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," *Database*, vol. 2020, Jan. 2020, Art. no. baaa010.

[3] J. L. Marcus, W. C. Sewell, L. B. Balzer, and D. S. Krakower, "Artificial intelligence and machine learning for HIV prevention: Emerging approaches to ending the epidemic," *Current HIV/AIDS Rep.*, vol. 17, no. 3, pp. 171–179, Jun. 2020.

[4] N. Dancy-Scott, G. A. Dutcher, A. Keselman, C. Hochstein, C. Copty, D. Ben-Senia, S. Rajan, M. G. Asencio, and J. J. Choi, "Trends in HIV terminology: Text mining and data visualization assessment of international AIDS conference abstracts over 25 Years," *JMIR Public Health Surveillance*, vol. 4, no. 2, p. e8552, Apr. 2018.

[5] A. Ovalle, O. Goldstein, M. Kachuee, E. S. C. Wu, C. Hong, I. W. Holloway, and M. Sarrafzadeh, "Leveraging social media activity and machine learning for HIV and substance abuse risk assessment: Development and validation study," *J. Med. Internet Res.*, vol. 23, no. 4, Apr. 2021, Art. no. e22042.

[6] C. Kakalou, J. V. Lazarus, and V. Koutkias, "Mining social media for 2090 perceptions and trends on HIV pre-exposure prophylaxis," *Stud. Health Technol. Informat.*, vol. 264, pp. 959–963, Aug. 2019.

[7] D. J. Feller, J. Zucker, M. T. Yin, P. Gordon, and N. Elhadad, "Using clinical notes and natural language processing for automated HIV risk assessment," *J. Acquired Immune Deficiency Syndromes*, vol. 77, no. 2, pp. 160–166, 2018.

[8] W.-T. Chen and R. Barbour, "Life priorities in the HIV-positive Asians: A text-mining analysis in young vs. old generation," *AIDS Care*, vol. 29, no. 4, pp. 507–510, 2017.

[9] HIV/AIDS. Mayo Clinic. (Mar. 26, 2022). *Mayo Foundation for Medical Education and Research*. Accessed: Jun. 5, 2022. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/hiv-aids/symptoms-causes/syc-20373524

[10] L. Fernandez-Luque and M. Imran, "Humanitarian health computing using artificial intelligence and social media: A narrative literature review," *Int. J. Med. Informat.*, vol. 114, pp. 136–142, Jun. 2018.

[11] A. Bahde, "Conceptual data visualization in archival finding aids: Preliminary user responses," *Portal, Libraries Acad.*, vol. 17, no. 3, pp. 485–506, 2017, doi: 10.1353/pla.2017.0031.

[12] B. Olatosi, X. Sun, S. Chen, J. Zhang, C. Liang, S. Weissman, and X. Li, "Application of machine-learning techniques in classification of HIV medical care status for people living with HIV in South Carolina," *AIDS*, vol. 35, no. 1, pp. S19–S28, May 2021.

[13] D. Fajardo-Ortiz, M. Lopez-Cervantes, L. Duran, M. Dumontier, M. Lara, H. Ochoa, and V. M. Castano, "The emergence and evolution of the research fronts in HIV/AIDS research," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0178293.

[14] D. G. Jamieson, M. Gerner, F. Sarafraz, G. Nenadic, and D. L. Robertson, "Towards semi-automated curation: Using text mining to recreate the HIV-1, human protein interaction database," *Database*, vol. 2012, Jan. 2012, Art. no. bas023.

[15] W.-T. Chen and R. Barbour. (2021). *Concerns Among The HIV-Positive Asians: A Data-Mining Analysis In Gender Comparison*. [Online]. Available: https://doi.org/10.21203/rs.3.rs-517685/v1

[16] T. Dhokotera, J. Bohlius, A. Spoerri, M. Egger, J. Ncayiyana, V. Olago, E. Singh, and M. Sengayi, "The burden of cancers associated with HIV in the south African public health sector, 2004–2014: A record linkage study," *Infectious Agents Cancer*, vol. 14, no. 1, pp. 2004–2014, Dec. 2019.

[17] P. Naveen, P. C. Nair, and D. Gupta, "Predicting the degree of emotional support in an online health forum for HIV using data mining techniques," in *Emerging Trends in Electrical, Communications, and Information Technologies*. Singapore: Springer, 2020, pp. 81–94.

[18] P. Breen, J. Kelly, T. Heckman, and S. Quinn, "Mining pre-exposure prophylaxis trends in social media," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics (DSAA)*, Oct. 2016, pp. 214–221.

[19] C. Liu and X. Lu, "Analyzing hidden populations online: Topic, emotion, and social network of HIV-related users in the largest Chinese online community," *BMC Med. Informat. Decis. Making*, vol. 18, no. 1, pp. 1–10, Dec. 2018, doi: 10.1186/s12911-017-0579-1.

[20] J. R. MacLean and K. Wetherall, "The association between HIV-stigma and depressive symptoms among people living with HIV/AIDS: A systematic review of studies conducted in south Africa," *J. Affect. Disorders*, vol. 287, pp. 125–137, May 2021.

[21] WHO. (2020). *World Health Organization. Global Health Observatory (GHO) Data: HIV/AIDS*. [Online]. Available: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/strategic-information/hiv-data-and-statistics

[22] O. Uwishema, C. Taylor, L. Lawal, N. Hamiidah, I. Robert, A. Nasir, and H. Onyeaka, "The syndemic burden of HIV/AIDS in Africa amidst the COVID-19 pandemic," *Immunity, Inflammation Disease*, vol. 10, no. 1, pp. 26–32, 2022.

[23] J. Dorward, T. Khubone, K. Gate, H. Ngobese, Y. Sookrajh, S. Mkhize, A. Jeewa, C. Bottomley, L. Lewis, K. Baisley, C. C. Butler, N. Gxagxisa, and N. Garrett, "The impact of the COVID-19 lockdown on HIV care in 65 South African primary care clinics: An interrupted time series analysis," *Lancet HIV*, vol. 8, no. 3, pp. e158–e165, Mar. 2021.

[24] M. Sallam, A. M. Alabbadi, S. Abdel-Razeq, K. Battah, L. Malkawi, M. A. Al-Abbadi, and A. Mahafzah, "HIV knowledge and stigmatizing attitude towards people living with HIV/AIDS among medical students in Jordan," *Int. J. Environ. Res. Public Health*, vol. 19, no. 2, p. 745, Jan. 2022.

[25] O. Oyebode, R. Lomotey, and R. Orji, "'I tried to breastfeed but' exploring factors influencing breastfeeding behaviours Based tweets using machine learning and thematic analysis," *IEEE Access*, vol. 9, pp. 61074–61089, 2021.

[26] A. A. Nichol, J. N. Batten, M. C. Halley, J. K. Axelrod, P. L. Sankar, and M. K. Cho, "A typology of existing machine learning-based predictive analytic tools focused on reducing costs and improving quality in health care: Systematic search and content analysis," *J. Med. Internet Res.*, vol. 23, no. 6, 2021, Art. no. e26391.

[27] L. I. C. C. Pinheiro, M. L. D. Pereira, M. P. Fernandez, F. M. V. Filho, W. J. C. P. de Abreu, and P. G. C. D. Pinheiro, "Application of data mining algorithms for dementia in people with HIV/AIDS," *Comput. Math. Methods Med.*, vol. 2021, pp. 1–8, Jul. 2021.

[28] Y. Xiao, C. Lin, L. Li, and G. Ji, "Individual and family level factors associated with physical and mental health-related quality of life among people living with HIV in rural China," *BMC Public Health*, vol. 19, no. 1, pp. 1–7, Dec. 2019.

[29] B. S. Anderson, "Using text mining to glean insights from COVID-19 literature," *J. Inf. Sci.*, vol. 49, no. 2, pp. 373–381, Apr. 2023.

[30] V. Cooper, J. Clatworthy, E. Youssef, C. Llewellyn, A. Miners, M. Lagarde, and M. Fisher, "Which aspects of health care are most valued by people living with HIV in high-income countries? A systematic review," *BMC Health Services Res.*, vol. 16, no. 1, pp. 1–15, 2016.

[31] Y. Nan and Y. Gao, "A machine learning method to monitor China's AIDS epidemics with data from Baidu trends," *PLoS ONE*, vol. 13, no. 7, Jul. 2018, Art. no. e0199697, doi: 10.1371/journal.pone.0199697.

[32] W. S. Comulada, C. Goldbeck, E. Almirol, H. J. Gunn, M. A. Ocasio, M. I. Fernández, E. M. Arnold, A. Romero-Espinoza, S. Urauchi, W. Ramos, M. J. Rotheram-Borus, J. D. Klausner, and D. Swendeman, "Using machine learning to predict young people's internet health and social service information seeking," *Prevention Sci.*, vol. 22, no. 8, pp. 1173–1184, Nov. 2021.

[33] C. Zheng, W. Wang, and S. D. Young, "Identifying HIV-related digital social influencers using an iterative deep learning approach," *AIDS*, vol. 35, no. 1, pp. S85–S89, May 2021.

[34] J. Shi, X. Wang, T.-Q. Peng, and L. Chen, "Understanding interactions in virtual HIV communities: A social network analysis approach," *AIDS Care*, vol. 29, no. 2, pp. 239–243, Feb. 2017, doi: 10.1080/09540121.2016.1210074.

[35] S. D. Young, W. Yu, and W. Wang, "Toward automating HIV identification: Machine learning for rapid identification of HIV-related social media data," *JAIDS J. Acquired Immune Deficiency Syndromes*, vol. 74, no. 2, pp. S128–S131, Feb. 2017.

[36] P. Garza, R. Sarvas, P. D. A. Malik, and A. Angi, "Applying natural language processing techniques to analyze HIV-related discussions on social media," M.S. thesis, Dept. Control Comput. Eng., Politecnico Di Torino, Italy, 2020.

[37] A. Thiabaud, I. Triulzi, E. Orel, K. Tal, and O. Keiser, "Social, behavioral, and cultural factors of HIV in malawi: Semi-automated systematic review," *J. Med. Internet Res.*, vol. 22, no. 8, Aug. 2020, Art. no. e18747.

[38] A. Hannaford, M. Lipshie-Williams, J. L. Starrels, J. H. Arnsten, J. Rizzuto, P. Cohen, D. Jacobs, and V. V. Patel, "The use of online posts to identify barriers to and facilitators of HIV pre-exposure prophylaxis (PrEP) among men who have sex with men: A comparison to a systematic review of the peer-reviewed literature," *AIDS Behav.*, vol. 22, no. 4, pp. 1080–1095, Apr. 2018.

[39] P. K. H. Mo and N. S. Coulson, "Exploring the communication of social support within virtual communities: A content analysis of messages posted to an online HIV/AIDS support group," *CyberPsychology Behav.*, vol. 11, no. 3, pp. 371–374, Jun. 2008.

[40] I. Hyndman, D. Nugent, G. G. Whitlock, A. McOwan, and N. Girometti, "COVID-19 restrictions and changing sexual behaviours in HIV-negative MSM at high risk of HIV infection in London, U.K," *Sexually Transmitted Infections*, vol. 97, no. 7, pp. 521–524, Nov. 2021.

[41] S. Oh and M. S. Park, "HIV/AIDS question analysis with text mining: Using concept maps for data analysis and interpretation," in *Proc. iConference*, 2015, pp. 1–4.

[42] Y. Ku, C. Chiu, B. H. Liou, J. H. Liou, and J. Y. Wu, "Applying text mining to assist people who inquire HIV/AIDS information from internet," in *Proc. Int. Conf. Intell. Secur. Inform.* Berlin, Germany: Springer, Jun. 2008, pp. 440–448.

[43] M. Mittal, I. Kaur, S. Pandey, A. Verma, and L. Goyal, "Opinion mining for the tweets in healthcare sector using fuzzy association rule," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 4, no. 16, Oct. 2018, Art. no. 159861, doi: 10.4108/eai.13-7-2018.159861.

[44] *Twitter*. [Online]. Available: https://twitter.com/

[45] T. Ye, R. Johnson, S. Fu, J. Copeny, B. Donnelly, A. Freeman, M. Lima, J. Walsh, and R. Ghani, "Using machine learning to help vulnerable tenants in New York city," in *Proc. 2nd ACM SIGCAS Conf. Comput. Sustain. Societies*, Jul. 2019, pp. 248–258.

[46] H. Lakkaraju, E. Aguiar, C. Shan, D. Miller, N. Bhanpuri, R. Ghani, and K. L. Addison, "A machine learning framework to identify students at risk of adverse academic outcomes," in *Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 1909–1918.

[47] N. Berberich, T. Nishida, and S. Suzuki, "Harmonizing artificial intelligence for social good," *Philosophy Technol.*, vol. 33, no. 4, pp. 613–638, Dec. 2020.

[48] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," *Health Inf. Sci. Syst.*, vol. 6, no. 1, pp. 1–12, Dec. 2018.

[49] V. S. Chavan and S. S. Shylaja, "Machine learning approach for detection of cyber-aggressive comments by peers on social media network," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Aug. 2015, pp. 2354–2358.

[50] A. Lillywhite and G. Wolbring, "Coverage of artificial intelligence and machine learning within academic literature, Canadian newspapers, and Twitter tweets: The case of disabled people," *Societies*, vol. 10, no. 1, p. 23, Feb. 2020.

[51] S. Alotaibi, R. Mehmood, I. Katib, O. Rana, and A. Albeshri, "Sehaa: A big data analytics tool for healthcare symptoms and diseases detection using Twitter, apache spark, and machine learning," *Appl. Sci.*, vol. 10, no. 4, p. 1398, Feb. 2020.

[52] S. R. Braithwaite, C. Giraud-Carrier, J. West, M. D. Barnes, and C. L. Hanson, "Validating machine learning algorithms for Twitter data against established measures of suicidality," *JMIR Mental Health*, vol. 3, no. 2, p. e4822, 2016.

[53] O. Oyebode and R. Orji, "Detecting factors responsible for diabetes prevalence in Nigeria using social media and machine learning," in *Proc. 15th Int. Conf. Netw. Service Manage. (CNSM)*, Oct. 2019, pp. 1–4.

[54] T. Sravanthi, V. Hema, S. T. Reddy, K. Mahender, and S. Venkateshwarlu, "Detection of mentally distressed social media profiles using machine learning techniques," *IOP Conf. Mater. Sci. Eng.*, vol. 981, no. 2, 2020, Art. no. 022056.

[55] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[56] Twitter. *Twitter API*. Accessed: May 2, 2022. [Online]. Available: https://developer.twitter.com/en/docs/twitter-api

[57] M. Konrad. (2017). Topic Model Evaluation in Python With TMTOOLKIT. WZB Data Science Blog. [Online]. Available: https://datascience.blog.wzb.eu/2017/11/09/topic-modeling-evaluation-in-python-with-tmtoolkit

[58] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Web Social Media*, 2014, vol. 8, no. 1, pp. 216–225.

[59] C. Sievert and K. Shirley, "LDAvis: A method for visualizing and interpreting topics," in *Proc. Workshop Interact. Lang. Learn., Visualizat., Interfaces*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014, pp. 1–8.

RICHARD K. LOMOTEY (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees. He is currently an IT Professor with The Pennsylvania State University, USA. He is also the Program Coordinator of the Information Sciences and Technology (IST) Program and the Cybersecurity Analytics and Operations (CYAOP) Program with Penn State Beaver. He is researching the intersection of smart technologies, mobile computing, data science, machine learning, and cybersecurity, with a focus on how these technologies are transforming enterprises and policies.

SANDRA KUMI (Graduate Student Member, IEEE) received the B.Sc. degree in computer science from the Kwame Nkrumah University of Science and Technology, Ghana, and the M.Sc. degree in computer science from Dongseo University, South Korea. She is currently pursuing the Ph.D. degree in computer science with the University of Saskatchewan, Canada. Her research interests include data trust, blockchain, cybersecurity, and machine learning.

MAXWELL HILTON is currently pursuing the Bachelor of Science degree with The Pennsylvania State University, PA, USA. He is also studying information science and technology, with a minor in security and risk analysis. His research interests include software development, machine learning, and cyber security.

RITA ORJI (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees. She is currently the Canada Research Chair of Persuasive Technology and a Professor in computer science with Dalhousie University, Canada, where she directs the Persuasive Computing Laboratory. Her research interests include the intersection of technology and human behavior, with a major focus on investigating user-centered approaches to designing technologies to improve lives and promote desirable behaviors. She applies her research to tackle real-life problems in various domains, including improving a wide range of health and wellness objectives, such as promoting mental health and discouraging risky health behaviors, including risky sexual behaviors, smoking cessation, promoting safety, security, and environmental sustainability.

RALPH DETERS (Member, IEEE) received the B.Sc. and M.Sc. degrees and the Ph.D. degree from Federal Armed Forces University, Munich, Germany, in 1998. He joined the University of Saskatchewan, as a Research Associate, in 1998, where he is currently a Full Professor with the Department of Computer Science. His research interests include distributed ledger technology, the IoT, and cloud/edge computing.

• • •