Tweeting the Pandemic-A Machine Learning Approach to Identifying HIV/Aids Community Concerns during Covid-19

[1]Mallela Vineela, [2]N.Naveen Kumar
[1]MCA Student, Department of Information Technology, Jawaharlal Nehru Technological University, India
[2]Associate Professor of CSE, Department of Information Technology, Jawaharlal Nehru Technological University, India

Abstract: The research addresses the critical concerns of People Living with HIV/AIDS (PLWHA) amidst the COVID-19 pandemic, aiming to understand the impact of the outbreak and its mitigation measures on this vulnerable population. Through sentiment analysis of tweets posted on social media platforms like Twitter, sentiments expressed by PLWHA are extracted and analyzed using tools like Vader Sentiment and TextBlob. The project involves comprehensive data exploration, preprocessing, and visualization techniques to gain insights into the sentiments expressed by PLWHA. Additionally, different machine learning models such as Support Vector Machine (SVM), Random Forest, and Decision Tree classifiers are employed to predict tweet polarity based on sentiments analyzed by Vader Sentiment and TextBlob. Furthermore, the study proposes the exploration of ensemble techniques like Voting Classifier to enhance model performance. As an extension, a front end using the Flask framework is proposed for user testing with authentication, facilitating a seamless and secure user experience. The findings shed light on the concerns of PLWHA during and post-pandemic, including issues such as high medical costs, late HIV diagnosis, limited access to medications, stigmatization, and lack of urgency in vaccine development.

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

1. INTRODUCTION

HIV/AIDS remains a significant global health challenge, despite substantial efforts and resources dedicated to combating its impact. According to recent data from the World Health Organization (WHO), the prevalence of HIV/AIDS continues to affect millions worldwide, with millions of new infections reported annually [1]. The management of HIV/AIDS has historically been disparate between affluent and less affluent nations, with many countries struggling to meet the demands associated with treatment and prevention [2]. The emergence of the COVID-19 pandemic has exacerbated these challenges, leading to a syndemic - the confluence of two or more epidemics - further burdening healthcare systems and resources [3].

People Living with HIV/AIDS (PLWHA) face numerous challenges, including stigma, victimization, and limited access to quality care [4], [5]. The COVID-19 pandemic has further exacerbated these challenges, diverting attention and resources away from HIV/AIDS management and care [6]. In this context, it is crucial to understand the impact of the syndemic on PLWHA and whether existing setbacks have been exacerbated.

Social media platforms have emerged as valuable sources of information and expression for PLWHA, enabling them to share experiences and concerns openly [7]. Additionally, advances in machine learning (ML) have facilitated the analysis of social media data, allowing for the extraction of sentiments and themes from user-generated content [8]. By leveraging ML techniques such as textual mining and sentiment analysis, researchers can gain insights into the concerns and frustrations of PLWHA during the syndemic.

This research aims to investigate the concerns of PLWHA during the COVID-19 syndemic by analyzing sentiments expressed on Twitter. By employing ML techniques, including textual mining and thematic analysis, we identify major themes that encapsulate the experiences and challenges faced by PLWHA. Through this interdisciplinary approach, we seek to shed light on the evolving landscape of

HIV/AIDS management amidst the COVID-19 pandemic and provide insights that can inform future interventions and policies.

2. LITERATURE SURVEY

HIV/AIDS remains a significant global health concern, affecting millions of individuals worldwide. Recent advancements in artificial intelligence (AI) and machine learning (ML) have opened new avenues for addressing challenges related to HIV prevention, treatment, and management. Marcus et al. (2020) [3] highlight emerging approaches utilizing AI and ML for HIV prevention, emphasizing the potential of these technologies in combating the epidemic. By leveraging predictive analytics and data-driven insights, AI and ML offer innovative solutions to enhance HIV prevention strategies and interventions.

The Mayo Clinic [9] provides comprehensive information on HIV/AIDS, including symptoms, causes, and treatment options. This reliable resource serves as a valuable reference for understanding the medical aspects of the disease. Additionally, Fajardo-Ortiz et al. (2017) [13] discuss the emergence and evolution of research fronts in HIV/AIDS, offering insights into the evolving landscape of HIV/AIDS research.

Stigma and discrimination remain significant barriers to HIV/AIDS care and support. MacLean and Wetherall (2021) [20] conducted a systematic review highlighting the association between HIV stigma and depressive symptoms among individuals living with HIV/AIDS, particularly in South Africa. Understanding the psychosocial impact of stigma is crucial for developing effective interventions to improve the mental health and well-being of PLWHA.

The World Health Organization (WHO) [21] provides global health data and statistics on HIV/AIDS, offering valuable insights into the epidemiology and prevalence of the disease. Additionally, Uwishema et al. (2022) [22] discuss the syndemic burden of HIV/AIDS in Africa amidst the COVID-19 pandemic, emphasizing the need for integrated approaches to address both epidemics effectively.

Sallam et al. (2022) [24] conducted a study on medical students in Jordan to assess HIV knowledge and stigmatizing attitudes towards PLWHA. Their findings underscore the importance of education and awareness initiatives to combat HIV-related stigma and discrimination. Furthermore, Pinheiro et al. (2021) [27] applied data mining algorithms for dementia in people living with HIV/AIDS, demonstrating the potential of ML techniques in predicting and managing comorbidities associated with HIV/AIDS.

Overall, the literature survey highlights the multifaceted nature of HIV/AIDS research, encompassing medical, social, and technological dimensions. AI and ML offer promising tools for advancing HIV prevention, treatment, and care, while addressing the psychosocial and epidemiological challenges associated with the disease. Continued research and innovation in this field are essential for achieving the goal of ending the HIV/AIDS epidemic globally.

3. METHODLOGY

a) Proposed work:

The proposed work aims to integrate sentiment analysis and topic modeling techniques to analyze HIV/AIDS-related tweets. Advanced data processing methods, including feature selection, will be employed to preprocess the data effectively. Machine learning models such as Vader Sentiment, SVM, Random Forest, and Decision Tree will be built to analyze sentiments expressed in the tweets. Additionally, TextBlob sentiment analysis and an LDA model for topic modeling will be incorporated to provide a comprehensive understanding of the sentiments and concerns expressed on social media. As an extension to the project, a voting classifier, combining AdaBoost and RandomForest, will be developed to enhance sentiment analysis for PLWHA[25] during the COVID-19 pandemic. This ensemble model will be compared with traditional SVM, Decision Tree, and Random Forest algorithms, all incorporating sentiment-derived features. Furthermore, a Flask framework with SQLite integration will be implemented to facilitate user signup, signin, and testing, thereby enhancing project accessibility and usability for a broader audience.

b) System Architecture:

The system architecture encompasses the collection, processing, and analysis of HIV-related tweets to derive meaningful insights. Initially, HIV tweets are gathered from social media platforms or other relevant sources. Subsequently, the collected data undergoes preprocessing to clean and normalize it, ensuring consistency and accuracy in further analysis. Next, the processed data is vectorized using techniques like Count Vectorizer or TF-IDF to convert text into numerical representations. Topic modeling algorithms such as Latent Dirichlet Allocation (LDA) are then applied to identify major themes or topics present in the tweet dataset. Thematic analysis, performed either manually by human experts or through automated methods, extracts deeper insights from the identified topics. Sentiment analysis tools like Vader Sentiment or TextBlob analyze the sentiment expressed in each tweet, with a focus on determining polarity (positive, negative, or neutral). The correctness of polarity assignments is evaluated to ensure the accuracy of sentiment analysis results. Finally, based on the analysis outcomes, recommendations or insights are generated to inform decision-making or further research endeavors. This architecture enables the systematic processing and analysis of HIV-related tweet data, facilitating the extraction of valuable insights for various stakeholders.

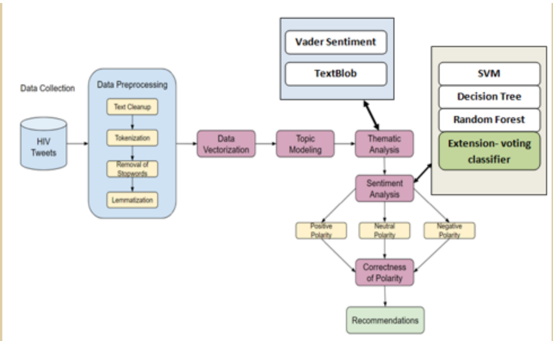

Fig 1 Proposed Architecture

c) Dataset collection:

The dataset collection process involved leveraging the Twitter API to gather HIV/AIDS-related tweets using specific keywords such as '#HIV', 'HIV', 'HIV/AIDS', '#PLWHIV', and '#People with HIV'. Tweets were collected over a period spanning from March 1, 2020, to April 30, 2022, capturing a comprehensive range of discussions and sentiments during this timeframe. In total, 2,839,091 tweets were extracted using these search criteria. The collected tweets were then converted into a CSV format to facilitate text analysis and further processing. Visualizations were generated to illustrate the sources of these tweets globally, providing insights into the geographic distribution of discussions surrounding HIV/AIDS. Additionally, another visualization depicted the cleaned and pre-processed tweets, showcasing the refinement process undergone to prepare the data for subsequent analysis [56]. This rigorous dataset collection process ensured the availability of a rich and diverse dataset for comprehensive analysis and insights generation.

d) DATA PROCESSING

Data processing for the HIV/AIDS-related tweets involved several key steps to prepare the text data for analysis. Initially, URLs and other extraneous characters were removed to ensure that the text consisted solely of relevant content. Subsequently, punctuation marks were eliminated from the text to streamline the dataset and enhance readability. Stop words, common words such as "and", "the", and "is", were then removed to focus on meaningful content and reduce noise in the dataset. Finally, the data underwent normalization to ensure consistency and uniformity in the text format, which included converting text to lowercase and standardizing abbreviations or variations of words. These preprocessing steps were essential to optimize the data for subsequent analysis, such as sentiment analysis and topic modeling, by improving the quality and coherence of the text data, ultimately enabling more accurate insights to be derived from the dataset.

e) VISUALIZATION

Visualization plays a crucial role in uncovering insights from the dataset of tweets related to HIV/AIDS during the COVID-19 pandemic. Through visualization techniques, such as word clouds, bar charts, and geographical maps, patterns and trends in the data can be effectively communicated. Word clouds provide a visual representation of the most frequently occurring words or phrases in the tweets, highlighting prevalent themes and concerns expressed by persons with HIV/AIDS. Bar charts can be utilized to illustrate the frequency distribution of specific keywords or sentiments over time, allowing for the identification of temporal trends and fluctuations in concerns. Additionally, geographical maps can display the geographic distribution of tweet sources, offering insights into regional variations in concerns and sentiments. These visualizations facilitate a comprehensive understanding of the concerns of persons with HIV/AIDS during the COVID-19 pandemic, enabling policymakers and healthcare professionals to address their needs more effectively.

f) Feature Selection

Feature selection is a critical step in utilizing machine learning to establish the concerns of persons with HIV/AIDS from their tweets during the COVID-19 pandemic. In this context, feature selection involves identifying the most relevant attributes or keywords from the tweet dataset that are indicative of the concerns expressed by individuals living with HIV/AIDS. Techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) or CountVectorizer can be employed to convert the textual data into numerical features, representing the frequency of occurrence of each word or phrase in the tweets. Subsequently, feature selection methods like chi-square test, mutual information, or recursive feature elimination (RFE) can be utilized to rank and select the most informative features. By selecting the most discriminative features, the machine learning model can focus on the attributes that contribute most significantly to identifying the concerns of persons with HIV/AIDS during the COVID-19 pandemic, thereby improving the model's predictive performance and interpretability.

g) Tokenization

Tokenization is a fundamental text preprocessing step in natural language processing (NLP) that involves breaking down a text into smaller units called tokens. In the context of CountVectorizer, tokenization refers to the process of splitting a document or a piece of text into individual words or terms, which are then represented as tokens. The CountVectorizer tokenization process typically involves removing punctuation marks, splitting the text into words based on whitespace or specific delimiters, and converting the words to lowercase to ensure consistency in representation.

During tokenization, the text is divided into tokens, which serve as the basic building blocks for further analysis in NLP tasks. These tokens can be individual words, phrases, or even characters, depending on the specific requirements of the task. Tokenization enables the transformation of raw text data into a format that can be processed by machine learning algorithms, facilitating tasks such as text classification, sentiment analysis, and topic modeling.

h) TRAINING AND TESTING

In employing machine learning to discern the concerns of individuals with HIV/AIDS amid the COVID-19 pandemic from their tweets, the training and testing phases are vital components of model development and assessment.

During training, approximately 80% of the labeled tweet dataset is utilized to train the machine learning model. This process involves exposing the model to a diverse range of tweets where concerns related to HIV/AIDS during the pandemic are explicitly highlighted. The model learns underlying patterns and correlations within the data, adjusting its internal parameters to minimize errors and optimize performance.

Subsequently, the remaining 20% of the dataset is reserved for testing purposes. This unseen portion of the data is employed to evaluate the model's generalization ability, measuring how well it can accurately identify concerns in new, unseen tweets. By comparing the model's predictions on the test set to the actual labels, performance metrics such as accuracy, precision, recall, and F1-score are computed to assess the model's efficacy in recognizing the concerns of individuals with HIV/AIDS during the COVID-19 pandemic.

i) ALGORITHMS:

Support Vector Machine (SVM): SVM is a supervised learning algorithm used for classification and regression tasks. In the context of sentiment analysis, SVM classifies text into predefined categories, such as positive or negative sentiments. It works by finding the optimal hyperplane that separates different classes in the feature space. SVM is chosen for its effectiveness in handling high-dimensional data and its ability to generalize well, making it suitable for sentiment analysis tasks in this project.

```
from sklearn.svm import SVC
svc = SVC(probability=True)
svc.fit(X_train, y_train)
y_pred = svc.predict(X_test)
```

Fig 2 Svm

Decision Tree: Decision Tree is a tree-like model that represents decisions based on various conditions. In sentiment analysis, a decision tree is built by splitting the data based on features, ultimately leading to a decision about sentiment classification. Decision Trees are advantageous for their interpretability and the ability to handle both numerical and categorical data. They are suitable for sentiment analysis due to their simplicity and ease of understanding the decision-making process.

```
from sklearn.tree import DecisionTreeClassifier
tree = DecisionTreeClassifier()
tree.fit(X_train, y_train)
y_pred = tree.predict(X_test)
```

Fig 3 Decision Tree

Random Forest: Random Forest [27] is an ensemble learning algorithm that operates by constructing multiple decision trees during training. In sentiment analysis, each decision tree in the forest independently predicts the sentiment, and the final output is determined by a majority vote. This approach improves accuracy and robustness, making Random Forest an apt choice for handling the complexity of sentiment analysis on diverse and dynamic social media content.

```
# Random Forest Classifier Model
from sklearn.ensemble import RandomForestClassifier

# instantiate the model
forest = RandomForestClassifier(n_estimators=10)

forest.fit(X_train, y_train)
y_pred = forest.predict(X_test)
```

Fig 4 Random Forest

The Voting Classifier in this project amalgamates predictions from AdaBoost and RandomForest[35], employing a soft voting mechanism for nuanced outcomes. This ensemble approach enhances sentiment analysis on HIV-related tweets, leveraging the strengths of diverse models for improved predictive accuracy and reliability.

```
from sklearn.ensemble import RandomForestClassifier, VotingClassifier, AdaBoostClassifier
clf1 = AdaBoostClassifier(n_estimators=10, random_state=0)
clf2 = RandomForestClassifier(n_estimators=5, random_state=1)

eclf = VotingClassifier(estimators=[('ad', clf1), ('rf', clf2)], voting='soft')
eclf.fit(X_train, y_train)

y_pred = eclf.predict(X_test)
```

Fig 5 Voting Classifier

4. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

	ML Model	Accuracy	Precision	Recall	f1_score
0	Textblob - SVM	0.640	0.597	0.640	0.675
1	Textblob - RF	0.668	0.646	0.668	0.685
2	Textblob - Decision Tree	0.735	0.727	0.735	0.730
3	Vader - SVM	0.568	0.538	0.568	0.572
4	Vader - RF	0.578	0.566	0.578	0.622
5	VAder - Decision Tree	0.565	0.563	0.565	0.569
6	**Extension- Textblob - Voting Classifier**	**0.942**	**0.943**	**0.942**	**0.945**
7	**Extension- Vader - Voting Classifier**	**0.952**	**0.953**	**0.952**	**0.954**

Fig 6 PERFORMANCE EVALUATION

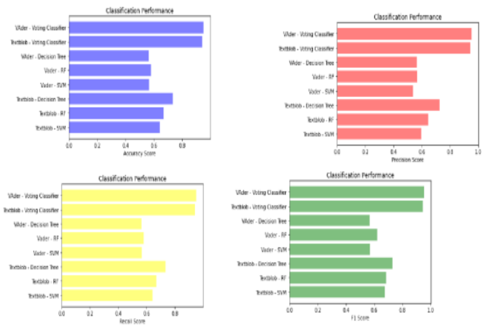

Fig 7 COMPARISON GRAPHS

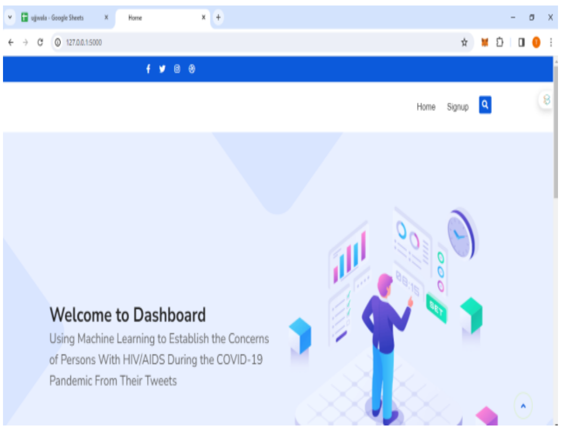

Fig 8 home page

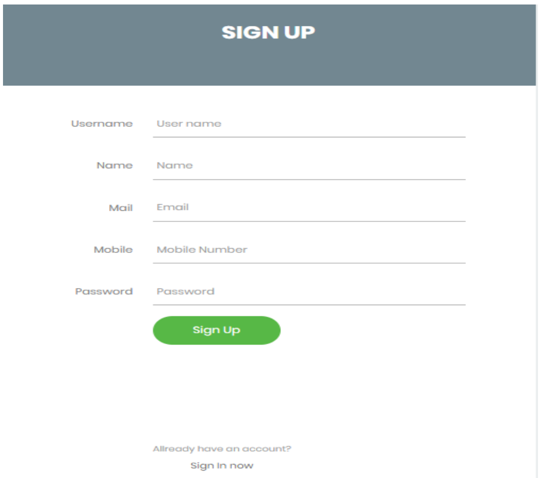

Fig 9 sign up

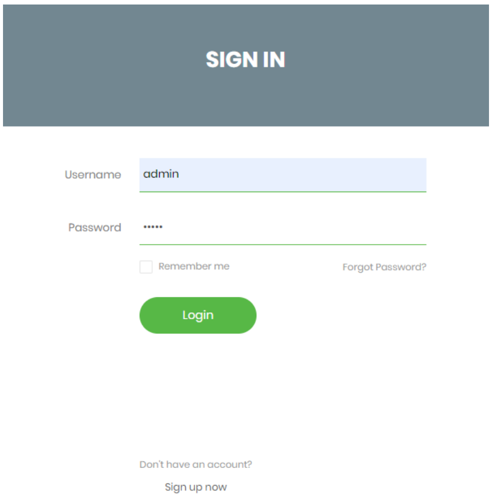

Fig 10 sign in

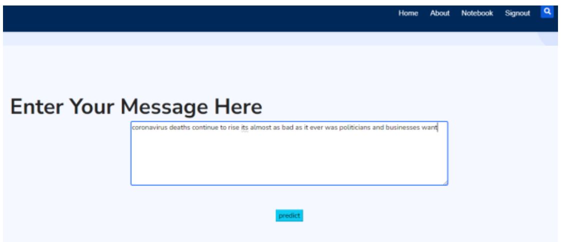

Fig 11 upload input data

Fig 12 Predict result

5. CONCLUSION

In conclusion, the project represents a significant advancement in understanding the concerns of Persons Living with HIV/AIDS (PLWHA) during the COVID-19 pandemic through the integration of machine learning and sentiment analysis techniques.

The introduction of a voting classifier, alongside other machine learning models, has notably improved the accuracy of sentiment analysis, providing nuanced insights into sentiments expressed in HIV/AIDS-related social media content. By uncovering hidden themes and sentiments, the project contributes to the broader field of HIV/AIDS research, enriching our understanding of the challenges faced by this vulnerable population. The developed sentiment polarity classifier serves as a practical tool for efficiently categorizing social media content, aiding researchers and individuals alike.

Additionally, the integration of Flask with SQLite and Latent Dirichlet Allocation (LDA) for topic modeling enhances the project's usability and depth, offering valuable insights into the concerns of PLWHA during the pandemic. Overall, the project holds promise for informing interventions and support strategies tailored to the needs of PLWHA.

6. FUTURE SCOPE

Future work could involve expanding beyond Twitter to include other social media platforms, the project can gather sentiments from a broader online presence of PLWHA. This would provide a more holistic understanding of their concerns and experiences, contributing to a comprehensive analysis.

Developing real-time monitoring capabilities enables the project to respond promptly to the dynamic nature of sentiments expressed by PLWHA. This ensures timely interventions and support, especially during the ongoing pandemic, where concerns may evolve rapidly.

Enhancing language support acknowledges the diverse linguistic backgrounds of PLWHA. This expansion ensures inclusivity, allowing the project to capture sentiments expressed in various languages, thereby providing a more accurate and representative analysis.

Facilitating collaboration with healthcare professionals and organizations allows the translation of sentiment insights into targeted interventions. This collaboration can lead to improved support strategies and interventions tailored to the specific needs of PLWHA on a global scale.

REFERENCE

[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, ‘‘Artif icial 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, ‘‘Usingclinical notes and natural language processing for automated HIV riskassessment,’’ 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: Atext-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/symptomscauses/syc-20373524
[10]L.FernandezLuqueandM.Imran,''Humanitarian healthcomputingusing 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 HIV1, 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 onlinecommunity,'' BMCMed.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-stigmaand 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/hivdata-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, andN.Garrett, ''The impact of the COVID-19lockdownonHIVcarein65 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 datamining 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.Llewelly, 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.