A Literature Review

AI-Based Surveillance of Depression in South Africa

Tristan Dos Remendos (2465830)

7 April 2024

Introduction

Depression, a pervasive mental health disorder, has garnered increasing attention in both historical discourse and contemporary medical research. Since ancient times, perceptions of depression have undergone significant transformations, reflecting evolving societal attitudes and advancements in medical understanding [28]. In recent decades, the recognition of depression as a leading cause of disability worldwide has underscored the urgency of effective disease surveillance and intervention strategies [35].

This literature review aims to explore the historical evolution of depression and its prevalence, and the methodologies employed in disease surveillance, particularly utilizing internet-based sources like Google Trends. By examining diverse approaches and contemporary research findings, this review seeks to shed light on the complex relationships between different factors in shaping the future of public health [27], as well as how that future can be improved with the aid of effective surveillance systems.

Central to this review is the exploration of internet-based sources as valuable tools for disease surveillance and monitoring. With the advent of technologies like Google Trends, researchers have gained new avenues to complement traditional surveillance approaches and enhance early warning capabilities [22]. Machine learning models, statistical regression analyses, and forecasting techniques offer promising opportunities to extract valuable insights from big data streams and inform public health interventions [30], [37].

The research problem/question guiding this review is thus twofold: firstly, to understand the statistical and health implications of the progression of depression over time, and secondly, to examine the challenges and opportunities presented by internet-based sources in health surveillance and monitoring, assessing its potential for being applied to a South African context. Through the exploration of these themes, this review aims to contribute to a comprehensive understanding of mental health surveillance and inform evidence-based interventions in a region that lacks sufficient research attention in the field of infodemiology [32], [35].

The structure of the review will proceed as follows: firstly, an exploration of the historical evolution of depression, followed by an examination of the prevalence of depressive disorders globally, and in South Africa. Subsequently, the focus will shift to methodologies employed in disease surveillance, with particular emphasis on internet-based sources. Finally, the review will conclude with a discussion of various statistical research methodologies commonly associated with public health surveillance and their efficacy when combined with internet-based sources.

1 A history of Depression

Depression has been recognised since ancient times, though the concepts and understanding of it have evolved considerably over history. Across various eras, depression saw changes in its perception ranging from a trait associated with genius and creativity, to imbalances of chemicals in the body. It was only after the 20th century, that the idea of depression became medicalised and treated as a serious health issue [28].

Around this time, the modern clinical concepts of depressive disorders began to take shape, with classifications developing in the Diagnostic and Statistical Manual (DSM) [38] and criteria for major depressive disorder, dysthymia, bipolar disorder and others emerging [28]. Biological treatments like antidepressant medications were also introduced during this period.

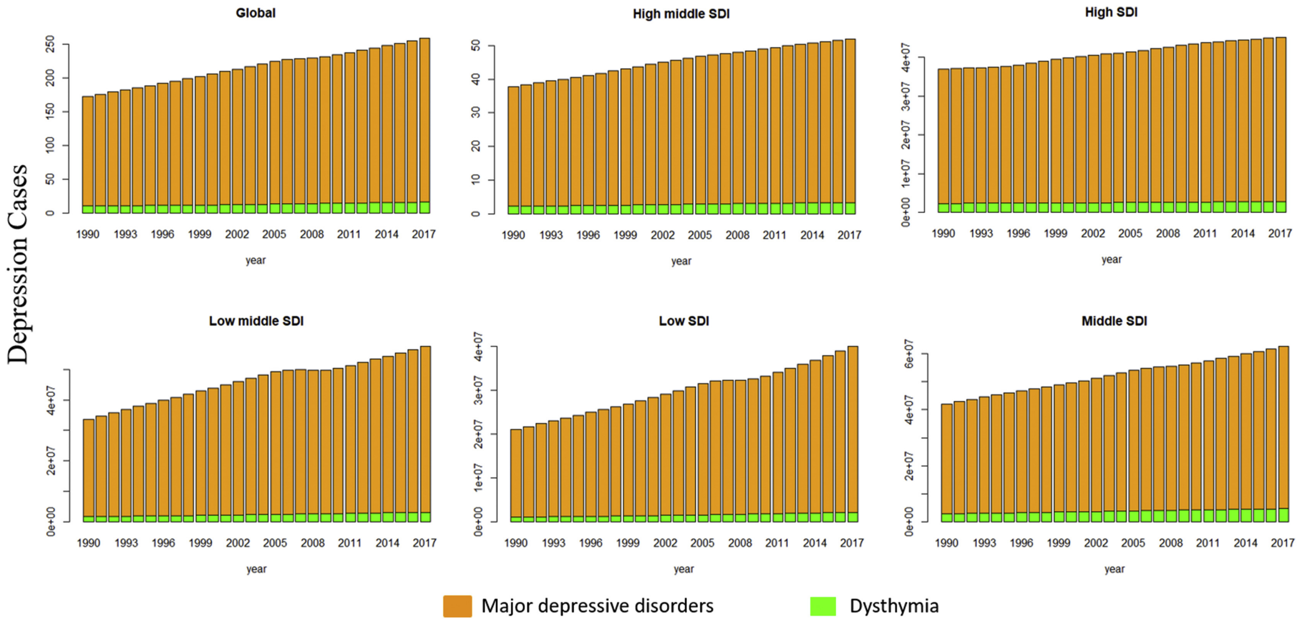

Figure 1: Depression cases, by SDI regions, from 1990 to 2017. [31]

1.1 Historical Statistics for Depression

Depression is a leading cause of disability worldwide, affecting people of all ages, backgrounds, and socioeconomic statuses. Its incidence continues to rise, as the World Health Organization (WHO) calculated approximately 280 million people suffered from depressive disorder in 2019 [35].

Apart from the significant impact of depression alone, mental health disorders collectively affect populations worldwide. According to the WHO, around 970 million people globally live with various mental health conditions [35]. This highlights the widespread prevalence and urgency for scientific study and targeted interventions to address the challenges posed by mental illness on a global level.

An article in the Journal of Psychiatric Research assessed the age-standardized incidence rate (ASR) of depression in over 190 countries [31]. The data used in the study was sourced from the Global Health Data Exchange, a website that provides a comprehensive database for all health-related data.

SDI refers to the sociodemographic index, which uses information on the economy, education, and fertility rate of different countries to represent their social/economic development. According to this study, it was observed that the incidence for depression cases in countries across different SDI regions is steadily increasing over time (see figure 1). This further states the importance of addressing depressive disorders as a health issue across the world.

1.2 Depression in South Africa

The history of depression and mental health in South Africa has been significantly shaped by the country's apartheid past and its socio-economic disparities. During the apartheid era, mental health services were largely segregated and inaccessible to the majority of the population, particularly in rural areas and for non-white communities [10].

After the end of apartheid in 1994, efforts were made to reform the mental health system and provide more equitable access to services. However, progress has been slow, and mental health remains a neglected area within the broader healthcare system [10], [25].

In terms of research on depression specifically, a study found that the proportion of South Africans who at one point were diagnosed with major depressive disorder was almost 10% [4]. More recently, the South African National Health and Nutrition Examination Survey (SANHANES-1) conducted in 2011-2012 reported a 4.5% prevalence of depression among South African adults [20].

However, it must be noted that the available research on depression in South Africa is still limited and can be relatively outdated. There is certainly a need for more comprehensive and ongoing surveillance to better understand the epidemiology and burden of depression across different regions, population groups, and socio-economic strata [10], [25].

2 The Surveillance of Disease

Disease epidemics have been of serious concern since historical times, much like the issues of depression mentioned before. An active measure towards the analysis and study of various diseases is that of disease surveillance, which entails the gathering, examination, and understanding of data regarding particular disease incidence within a population [2].

According to Centres for Disease Control (CDC), public health surveillance plays a vital role in detecting disease outbreaks early, monitoring trends, assessing intervention effectiveness, allocating resources efficiently, and informing policy development [1]. By systematically collecting, analyzing, and disseminating data on specific health events within a population, surveillance enables timely responses to emerging threats, evaluation of control measures, and evidence-based decision-making. The CDC believes this continuous monitoring helps protect public health by identifying and addressing health challenges effectively.

Over time, there have been different approaches to disease surveillance, and advancements in technology have enhanced our ability to statistically monitor and analyze data not only for diseases but also for other health incidents [30]. Each new approach yields different benefits and limitations when being compared to one another.

2.1 Traditional Approaches

Traditional disease-specific surveillance entails monitoring diseases through laboratory testing and subsequent reporting. Public health departments analyze the data from these reports to detect trends and patterns in disease occurrence and transmission [37]. Laboratory testing offers comprehensive diagnosis and identifies the particular strain of a pathogen, facilitating the tracking of global trends in surveyed pathogens and infectious diseases [18].

Beyond traditional disease-specific surveillance, we can draw parallels to the surveillance of depression. Similarly, surveillance of depression involves monitoring individuals' mental health through diagnostic assessments rather than laboratory testing. Public health departments analyze data from clinical evaluations and self-report measures to identify trends and patterns in the occurrence and prevalence of depression within populations [26]. While diagnoses are made based on patients' symptoms and behaviours rather than biological samples, the overall objective remains the same: to understand and address the burden of health issues.

However, this approach is not without its limitations. One overarching drawback of the traditional method is the significant delay between diagnosing a disease or disorder and reporting it. This delay, often referred to as a "time lag" can vary from days to weeks, hindering the timely implementation of interventions [37]. Moreover, studies have shown substantial diagnosis delays for certain conditions. For example, research conducted in Spain found that the average delay in diagnosing major depressive disorder was nearly 10 weeks [13]. Consequently, the combined time required for physical diagnosis of depressive disorders and the subsequent reporting delay pose significant challenges in promptly treating patients with depressive disorders.

Additionally, the resource demands associated with traditional surveillance methods can be burdensome, particularly for lower-income countries. Moreover, these methods may result in inaccuracies in reported information due to delays in data collection and reporting [30]. Early intervention and diagnosis are crucial in alleviating the burden of mental disorders, underscoring the paramount importance of timely surveillance data [13].

2.2 New Ways to Survey Diseases

Based on the aforementioned limitations associated with traditional surveillance methods, there is a clear space for the modernisation of these methods. The first instances of this modernisation occurred as a result of the revelation of the internet. The introduction of internet-based sources (IBS) has since facilitated the development of new methodologies for monitoring and tracking the spread of diseases [27]. Examples of these sources include search queries, social media platforms and web-sites [7], [14], [16].

The use of online data streams for public health monitoring, termed infodemiology or infoveillance, took root in the early 2000s with systems like ProMED-mail harnessing sources like emails and web feeds to detect potential outbreaks [3]. A major milestone was the 2009 study using Google search data to estimate flu activity [7], paving the way for Google's launch of Flu Trends before then end of of the decade. Around 2010, microblogs like Twitter garnered research interest for containing health mentions, with efforts to identify true personal health statements indicating infection [8].

Over time, the scope expanded from early focuses on flu tracking and infectious diseases to also monitoring chronic illnesses [21], leveraging an increasing array of online sources like social media [16], forums [17], news media [19] and Wikipedia [14]. Methodologies progressed from simple correlational analyses to deploying machine learning and deep learning models on the unstructured, big data streams [27]. While search query logistics powered early tools, the field evolved to tap the wealth of contextualized health information shared by users across diverse online platforms.

2.2.1 Types of Health Issues that have been Surveyed using New Approaches

Studies utilizing online data sources for disease surveillance have covered a wide range of infectious and chronic diseases. Examples include influenza tracking using search engine query data [7], detection of influenza-like illness through analysis of blog post frequencies [6], monitoring of sexually transmitted infections using Google Trends data [30], and surveillance of HIV using social media data [29]. A number of studies used online data methods to survey or forecast Covid-19 cases [33], [34].

However, there remains a scarcity of infodemiology studies focusing specifically on depression. Some studies include: population mapping of depression in the US [36] and temporal variation of depression statistics in Finland [24]. Another study looked at the association of searches with possible suicide deaths [11], which is a nuanced avenue for connecting depression trends to online data. Nevertheless, further research is needed to expand the application of infodemiology to depression and other mental health conditions in various world regions, especially Africa.

2.2.2 Google Trends as a Digital Surveillance Tool

The general method for the use of Google Trends (GT) for topical surveillance involves retrieving data from Google Trends, which represents the relative search volume for specific keywords or topics over time. Researchers typically select relevant keywords or phrases related to the health condition or event of interest, such as disease names, symptoms, or related terms [11], [30], [34], [37]. The search volume data is then analyzed, often in combination with other data sources (e.g., official health statistics, demographic data), to identify patterns, trends, and potential correlations with actual health events or outcomes.

Benefits:

- Timeliness: Google Trends data is available in near real-time, allowing for early detection and monitoring of health events, potentially before official reports [11], [30].
- Cost-effective: Utilizing freely available search data can be a cost-effective alternative or complement to traditional surveillance methods [34], [37].
- Geographical coverage: Google Trends data can provide insights into health trends across different regions or countries, including areas with limited traditional surveillance systems [34].

Limitations:

- Representativeness: The user demographics and internet access patterns of Google users may not accurately represent the general population, potentially introducing biases [11], [37].
- Noise and confounding factors: Search volume can be influenced by various factors, such as media coverage, seasonality, or unrelated events, making it challenging to isolate the signal related to the health condition of interest [11], [30].
- Lack of specificity: Search queries may be ambiguous or lack sufficient context to accurately identify the underlying health condition or event [30], [37].

While a number of studies that utilize Google Trends for the surveillance of disease report good results for correlation [37] (see figure 2), and some even demonstrate the ability to forecast/predict the trends into the future to some degree [30]. There are studies that have reported that Google Trends may be difficult to

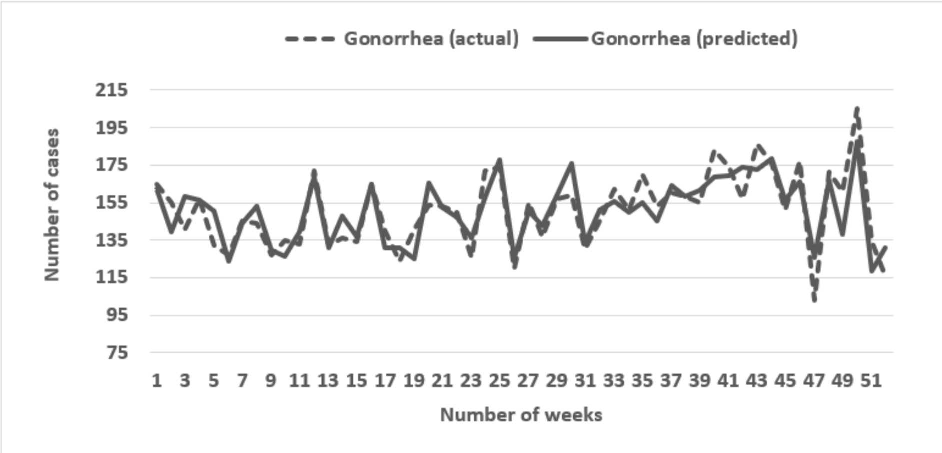

Figure 2: Comparison between actual and predicted number of gonorrhea cases for 2017, conducted in United States (Chicago), with the use of GT [30].

implement as a tool for surveillance.

A study utilized Google Trends to evaluate Covid-19 incidence in Africa, concluding that diseases with seasonal patterns may be more effectively surveyed using Google Trends compared to the unprecedented Covid-19 epidemic [34]. Factors such as misreporting of Covid-19 cases and the adverse effects of "social media influence" are believed to have contributed to the relatively limited success of Google Trends in Covid-19 surveillance.

Additionally, the efficacy of Google Trends in indicating disease incidence may be hindered by language differences in certain countries, making it challenging to select search terms accurately associated with the diseases [34].

On the flip side, several studies have highlighted successes in identifying correlations or utilizing Google Trends as a dependable data reservoir for statistical examination. For instance, researchers exploring the mapping of depression prevalence across the United States advocated for the use of Google Trends data as an innovative digital epidemiological tool. They argued that such data can efficiently map depression prevalence and trends within populations, offering real-time insights at a low cost and overcoming constraints associated with traditional surveys [36].

The sentiment surrounding online big data for public health surveillance suggests that analyzing information from search engines can complement surveys, enhancing our understanding of the burden of mental health conditions, provided that we acknowledge and address potential limitations in data sources [27]. Any research into the use of these types of data sources are beneficial towards the improvement of current schemes for public health surveillance.

2.3 Geographic Coverage of the Utility of Online Sources for Public Health Surveillance

The review conducted in 2014 assessed the use of IBS frameworks for monitoring diseases across various global regions [15]. While the review highlighted the potential for IBS tools to provide global coverage for disease monitoring, it also pointed out disparities in their effectiveness and implementation across different countries and regions.

One significant finding is that IBS systems are ideal for large populations, but the implementation depends on resource constraints (or infastructure) [5]. This limitation is especially present in low-income countries with lower rates of internet access. For example, the percentage of the population which can be evaluated by an IBS framework is much higher in high-income countries with high internet access (77% internet users) compared to low-income countries (31% internet users) in 2013 [12] (see figure 3).

Furthermore, cultural, linguistic, and behavioral differences in how people seek and share health information online can also hinder the implementation and performance of IBS systems across different countries and regions [9]. These differences may introduce biases that affect the accuracy of the surveillance data collected.

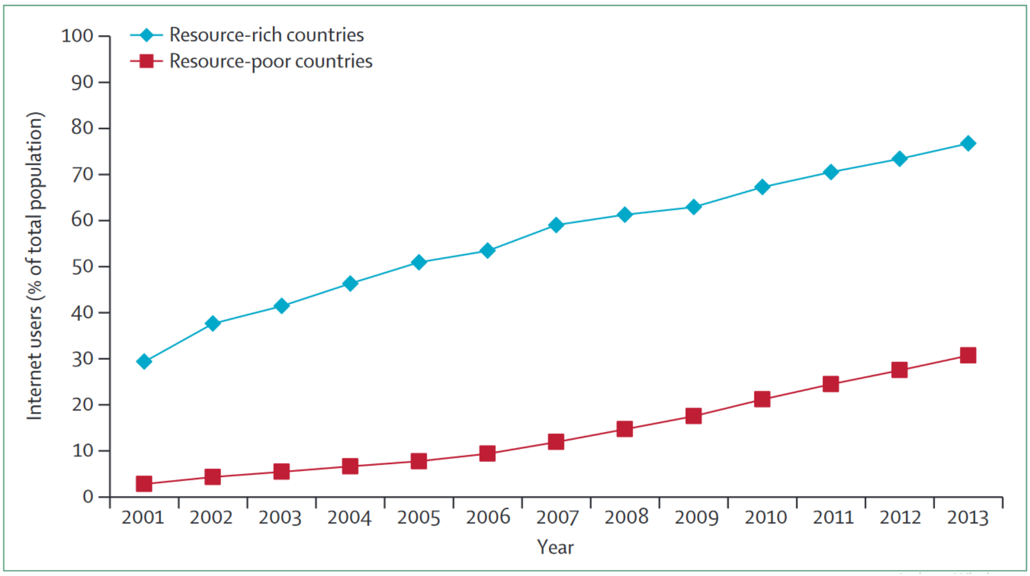

Figure 3: Internet usage statistics in low vs. high income countries [15].

For high-income regions, such as North America and Europe, there are a large number of studies specifically assessing/utilizing IBS for public health surveillance. Some countries include: United States [14], [16], [30], [36], Canada [21], Portugal [37] and England [7].

In the context of Africa, the challenges mentioned are particularly pronounced. Many countries in Africa face infrastructural challenges, including limited internet access and connectivity. Additionally, cultural and linguistic diversity across the continent can further complicate the implementation of IBS systems for health surveillance. As a result, there is a scarcity of studies documenting the use of IBS surveillance in Africa as indicated by a paper released in 2020 [32], highlighting the need for tailored approaches and strategies to overcome these barriers and leverage the potential of online surveillance for monitoring health in the region.

As previously mentioned, a study utilized IBS investigated Google Trends data to measure COVID cases in Africa [34]. Another study used news reports to characterize epidemiological trends for Ebola in countries in West-Africa [19].

However, based on my research findings, there appears to be only a handful of studies regarding the utilization of IBS surveillance systems/tools specifically in South Africa. A study found utilized Google Trends data to forecast influenza-like illness [33], as well as a study that used social media data for surveillance of HIV [29]. Ultimately, it is clear that South Africa in particular is under-researched in terms of its potential for big-data based health surveillance.

3 Statistical Research Methodologies

Most studies that were reviewed employed a variety of statistical research methodologies to analyze internet-based data, particularly from Google Trends, for disease surveillance and monitoring. These can be broadly categorized into machine learning models, statistical regression models, and approaches comparing correlation versus forecasting.

Machine learning techniques and regression analysis offer distinct advantages in disease and public health surveillance [22]. Machine learning excels in handling large and complex datasets with nonlinear relationships, providing high prediction accuracy, automating feature selection, and scaling efficiently to big data. On the other hand, regression analysis offers a more interpretable framework for analyzing relationships between variables, conducting hypothesis testing, and estimating the effect size of covariates on disease outcomes [22]. Both approaches play essential roles in advancing our understanding of disease dynamics and informing public health interventions.

Several studies utilized advanced machine learning techniques to develop predictive models from Google

Trends data. For example, [37] explored different machine learning algorithms, including Partial Least Squares (PLS), Ordinary Least Squares (OLS), LASSO, Support Vector Machines, and Deep Neural Networks, to forecast monthly tuberculosis incidence in Portugal. The author found that the PLS model achieved the best predictive performance across various evaluation metrics.

Similarly, [30] applied elastic net regression modeling using Google Trends data as predictors to forecast sexually transmitted infection (STI) case counts in Chicago. These machine learning-based approaches demonstrate the potential to leverage the rich information contained in search query data to enhance disease monitoring and early warning capabilities.

In addition to machine learning, several studies employed more traditional statistical regression techniques. [34] and [36] both utilized regression analyses to assess the correlation between Google Trends search volumes and reported disease incidence or prevalence.

[34] found very weak correlations between Google Trends data and COVID cases across African populations, suggesting limited applicability of this approach for emerging epidemics. In contrast, [36] reported stronger associations between Google Trends data and depression prevalence in the United States, highlighting the potential for this method to complement traditional epidemiological surveys.

3.1 Correlation vs Forecasting

Based on my findings, while some studies focused on assessing the correlation between internet-based data and official disease statistics, others explored the use of these data sources for forecasting and early detection.

Correlation and forecasting are two distinct statistical techniques used in data analysis. Correlation measures the strength and direction of association between two variables, indicating how they change together without implying causation [23]. On the other hand, forecasting involves using historical data to make predictions about future values of a variable, aiming to estimate future behavior based on past observations and trends [23]. While correlation helps identify potential relationships between variables, forecasting is useful for anticipating future trends and making informed decisions based on anticipated outcomes.

The work by [6] demonstrated a high correlation ($r = 0.626$) between influenza-related blog post frequencies and CDC influenza-like illness surveillance data. This suggested the potential utility of web-based data for augmenting traditional surveillance approaches.

In comparison, the studies by [30] and [37] went beyond correlation to develop predictive models that could "nowcast" or forecast near-real-time disease trends, potentially enabling earlier public health interventions. This forecasting capability is a key advantage of leveraging internet-based data sources for infodemiology, however, is also a more time-consuming approach due to the requirements of model training and evaluation.

Conclusion

In conclusion, this literature review has explored the historical evolution of depression and its prevalence, and the methodologies employed in disease surveillance, particularly focusing on internet-based sources like Google Trends. Through an examination of historical and contemporary research findings, several key themes have emerged.

The global prevalence of mental health disorders underscores the urgency of effective disease surveillance and intervention strategies [35]. Some regions are better equipped for enabling newer or augmented surveillance techniques but the usefulness of any research in the use of these techniques for under-equipped, less data-dense regions cannot be emphasized further [15].

Importantly, the exploration of internet-based sources as valuable tools for disease surveillance and monitoring yielded positive discoveries about the potential applications of these tools to depression surveillance. Machine learning models, statistical regression analyses, and forecasting techniques offer promising opportunities to complement traditional surveillance approaches and enhance early warning capabilities [22], [29], [37].

In summary, this literature review provides valuable insights into the complexities of health surveillance and highlights the potential of internet-based sources to enhance disease monitoring and intervention efforts, with prospects of the application of these sources and models to the infodemiology of depression in South African contexts.

References

[1] A. Friede, J. A. Reid, and H. W. Ory, "Cdc wonder: A comprehensive on-line public health information system of the centers for disease control and prevention.," *American Journal of Public Health*, vol. 83, no. 9, pp. 1289–1294, 1993.

[2] S. B. Thacker, D. F. Stroup, R. G. Parrish, and H. A. Anderson, "Surveillance in environmental public health: Issues, systems, and sources.," *American journal of public health*, vol. 86, no. 5, pp. 633–638, 1996.

[3] V. L. Yu and L. C. Madoff, "ProMED-mail: An Early Warning System for Emerging Diseases," *Clinical Infectious Diseases*, vol. 39, no. 2, pp. 227–232, Jul. 2004. [Online]. Available: https://doi.org/10.1086/422003.

[4] D. R. Williams, A. Herman, D. J. Stein, *et al.*, "Twelve-month mental disorders in south africa: Prevalence, service use and demographic correlates in the population-based south african stress and health study," *Psychological Medicine*, vol. 38, no. 2, pp. 211–220, 2008.

[5] H. A. Carneiro and E. Mylonakis, "Google trends: A web-based tool for real-time surveillance of disease outbreaks," *Clinical infectious diseases*, vol. 49, no. 10, pp. 1557–1564, 2009.

[6] C. D. Corley, D. J. Cook, A. R. Mikler, and K. P. Singh, "Using web and social media for influenza surveillance," in *Advances in Experimental Medicine and Biology*, vol. 680, 2009, pp. 559–564.

[7] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, no. 7232, pp. 1012–1014, 2009.

[8] A. Culotta, "Towards detecting influenza epidemics by analyzing twitter messages," in *Proceedings of the First Workshop on Social Media Analytics*, ser. SOMA '10, Washington D.C., District of Columbia: Association for Computing Machinery, 2010, pp. 115–122, ISBN: 9781450302173. [Online]. Available: https://doi.org/10.1145/1964858.1964874.

[9] T. M. Hale, S. R. Cotten, P. Drentea, and M. Goldner, "Rural-urban differences in general and health-related internet use," *American Behavioral Scientist*, vol. 53, no. 9, pp. 1304–1325, 2010.

[10] C. Lund, S. Kleintjes, R. Kakuma, and A. J. Flisher, "Public sector mental health systems in south africa: Inter-provincial comparisons and policy implications," *Social Psychiatry and Psychiatric Epidemiology*, vol. 45, no. 3, pp. 393–404, 2010.

[11] A. C. Yang, S.-J. Tsai, N. E. Huang, and C.-K. Peng, "Association of internet search trends with suicide death in taipei city, taiwan, 2004–2009," *Journal of Affective Disorders Reports*, 2011.

[12] J. Watson, R. Pebody, *et al.*, "Influenza surveillance and pandemic requirements.," *Pandemic influenza*, no. Ed. 2, pp. 9–18, 2012.

[13] R. Huerta-Ramírez, J. Bertsch, M. Cabello, M. Roca, J. M. Haro, and J. L. Ayuso-Mateos, "Diagnosis delay in first episodes of major depression: A study of primary care patients in spain," *Journal of affective disorders*, vol. 150, no. 3, pp. 1247–1250, 2013.

[14] D. J. McIver and J. S. Brownstein, "Wikipedia usage estimates prevalence of influenza-like illness in the united states in near real-time," *PLoS computational biology*, vol. 10, no. 4, e1003581, 2014.

[15] G. J. Milinovich, G. M. Williams, A. C. Clements, and W. Hu, "Internet-based surveillance systems for monitoring emerging infectious diseases," *The Lancet infectious diseases*, vol. 14, no. 2, pp. 160–168, 2014.

[16] S. Gittelman, V. Lange, C. A. Gotway Crawford, *et al.*, "A new source of data for public health surveillance: Facebook likes," *Journal of medical Internet research*, vol. 17, no. 4, e98, 2015.

[17] C. Karmen, R. C. Hsiung, and T. Wetter, "Screening internet forum participants for depression symptoms by assembling and enhancing multiple nlp methods," *Computer Methods and Programs in Biomedicine*, vol. 120, no. 1, pp. 27–36, 2015, ISSN: 0169-2607. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260715000620.

[18] C. Abat, H. Chaudet, J.-M. Rolain, P. Colson, and D. Raoult, "Traditional and syndromic surveillance of infectious diseases and pathogens," *International Journal of Infectious Diseases*, vol. 48, pp. 22–28, 2016.

[19] J. M. Cleaton, C. Viboud, L. Simonsen, A. M. Hurtado, and G. Chowell, "Characterizing ebola transmission patterns based on internet news reports," *Clinical Infectious Diseases*, vol. 62, no. 1, pp. 24–31, 2016.

[20] V. Pillay-van Wyk, W. Msemburi, R. Laubscher, *et al.*, "Mortality trends and differentials in south africa from 1997 to 2012: Second national burden of disease study," *The Lancet Global Health*, vol. 4, no. 9, e642–e653, 2016.

[21] W. Zhang, S. Ram, M. Burkart, and Y. Pengetnze, "Extracting signals from social media for chronic disease surveillance," in *Proceedings of the 6th International Conference on Digital Health Conference*, ser. DH '16, Montréal, Québec, Canada: Association for Computing Machinery, 2016, pp. 79–83, ISBN: 9781450342247. [Online]. Available: `https://doi.org/10.1145/2896338.2897728`.

[22] D. Bzdok and N. Altman, *Machine Learning: A Guide to Current Research*. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2018, vol. 12, pp. 1–207. DOI: `10.2200/s00868ed1v01y201810aim040`.

[23] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[24] J. Tana, "An infodemiological study using search engine query data to explore the temporal variations of depression in finland," *Finnish Journal of eHealth and eWelfare*, vol. 10, no. 1, pp. 133–142, 2018.

[25] S. Docrat, C. Lund, and D. Chisholm, "Sustainable financing options for mental health care in south africa: Findings from a situational analysis and key informant interviews," *International Journal of Mental Health Systems*, vol. 13, no. 1, pp. 1–14, 2019.

[26] C. Lang, "Inspecting mental health: Depression, surveillance and care in kerala, south india," *Culture, Medicine, and Psychiatry*, vol. 43, no. 4, pp. 596–612, 2019.

[27] J. M. Barros, J. Duggan, and D. Rebholz-Schuhmann, "The application of internet-based sources for public health surveillance (infoveillance): Systematic review," *Journal of medical internet research*, vol. 22, no. 3, e13680, 2020.

[28] M. Bourin, "History of depression through the ages," *Archives of Depression and Anxiety*, May 2020.

[29] A. van Heerden and S. Young, "Use of social media big data as a novel hiv surveillance tool in south africa," *Plos one*, vol. 15, no. 10, e0239304, 2020.

[30] A. K. Johnson, R. Bhaumik, I. Tabidze, and S. D. Mehta, "Nowcasting sexually transmitted infections in chicago: Predictive modeling and evaluation study using google trends," *JMIR PUBLIC HEALTH AND SURVEILLANCE*, 2020.

[31] Q. Liu, H. He, J. Yang, X. Feng, F. Zhao, and J. Lyu, "Changes in the global burden of depression from 1990 to 2017: Findings from the global burden of disease study," *Journal of Psychiatric Research*, vol. 126, pp. 134–140, 2020. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0022395619307381`.

[32] S. Olukanmi and F. Nelwamondo, "Digital influenza surveillance: The prospects of google trends data for south africa," in *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, IEEE, 2020, pp. 1–6.

[33] S. O. Olukanmi, F. V. Nelwamondo, and N. I. Nwulu, "Utilizing google search data with deep learning, machine learning and time series modeling to forecast influenza-like illnesses in south africa," *IEEE Access*, vol. 9, pp. 126 822–126 836, 2021.

[34] A. Fulk, D. Romero-Alvarez, Q. Abu-Saymeh, J. S. Onge, A. Peterson, and F. Agusto, "Using google health trends to investigate covid-19 incidence in africa," *PLOS One*, Jun. 2022.

[35] S. Lewis, M. Freeman, M. van Ommeren, D. Chisholm, O. G. Siegl, and D. Kestel, "World mental health report," World Health Organisation, Tech. Rep., 2022.

[36] A. Wang, R. McCarron, D. Azzam, A. Stehli, G. Xiong, and J. DeMartini, "Utilizing big data from google trends to map population depression in the united states: Exploratory infodemiology study," University of California, Tech. Rep., 2022.

[37] L. G. Santos, "Surveillance of tuberculosis by analysing google trends," *Católica Lisbon School of Economics and Business*, Jun. 2023.

[38] A. P. Association, *The Diagnostic and Statistical Manual of Mental Disorders*. Originally published in 1952, latest revision in 2022, p. 947, ISBN: 978-0-89042-554-1.