Procedia of Social Sciences and Humanities
International Conference On Emerging New Media and Social Science

Act or Wait and See the Challenges: Artificial Intelligence for Analysing Schizophrenia Syndromes in Social Media

Akash Gulati[1], Dr. Sugandh Arora[2], Dr. Kirti Dang Longani[3*]
[1] ICFAI Business School Hyderabad, India
[2,3]Ajeenkya DY Patil University, India
*Correspondence Author Email: kirtidanglongani@gmail.com

Abstract. *Schizophrenia is a complicated and crippling mental illness that makes it hard to find and treat early on. With the rise of social media, there is a lot of information available to help people learn more about mental health issues, such as schizophrenia syndromes. It might be possible to find possible signs of schizophrenia and improve the diagnostic process by using artificial intelligence (AI) techniques to look at social media data. The frequent use of social media can be indicative of linguistic impairments or alterations brought on by symptoms shared by a variety of mental health illnesses. Over the past 25 years, the detection of these linguistic cues has been studied; however, with the pandemic, interest and methodological advancement have increased dramatically. It is possible that within the next ten years, trustworthy techniques for utilising social media data to forecast mental health status will emerge. This could have an impact on public health policy and clinical practise, especially when it comes to early intervention in mental health treatment.*

Keywords: *Artificial Intelligence (AI), Schizophrenia Syndromes, Social Media*

I. INTRODUCTION

About 20 million people worldwide suffer from schizophrenia, a complex psychiatric illness that causes disruptions in thought, perception, emotions, and behavior. Schizophrenia places a heavy strain on sufferers, families, and the healthcare system. Schizophrenia is a disorder marked by a variety of symptoms, such as delusions, social disengagement, disorganized thinking, and hallucinations. Early diagnosis, treatment, and detection of schizophrenia can be extremely difficult. Social media's rapid growth in popularity has changed the way people engage, connect, and express themselves online in recent years. Social media, with billions of users worldwide utilizing a variety of digital communication platforms, has developed into a priceless real-time data source of user-generated information representing a range of human behavior features, including mental health. On social media, those impacted by serious mental health illnesses frequently share their experiences . In general, people use text on Facebook and Reddit to communicate both positive and negative life events, while image-focused platforms like Instagram are seeing an increase in the usage of pictures to communicate delicate subjects like illness or hardship . Consequently, there has been a surge in techniques that use social media data to forecast users' mental health state (Zhang, 2021). Following the COVID-19 epidemic , research also increased significantly and evolved into a fully multidisciplinary field comprising not only computer scientists but also psychologists, psychiatrists, and neuroscientists. This field's overarching theory is that models powered by artificial intelligence (AI) have the ability to "predict" an individual's "mental health condition" (see for a discussion on the meaning of these terms in this literature). Natural Language Processing (NLP), a subfield of artificial intelligence that uses computational methods to

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

learn, comprehend, and synthesize human language material, is best suited for these approaches. The vast amount of online human language content, such as posts on Facebook and Reddit, easily fits the technological limitations of natural language processing (NLP) approaches and may be easily converted into model inputs. Using traditional statistical analyses and manual posting review, some of the first attempts at forecasting the mental health statuses of members of online communities were carried out before artificial intelligence was ever invented. In November 1999, for instance, psychiatrists kept an eye on the general psychiatry subforum in the Norwegian web forum Doktoronline. They noticed that people who expressed despair or resignation in their writings on their mental health usually got supportive and helpful feedback from other users. Affected users subsequently frequently looked for social assistance in their neighborhoods. This validated earlier research demonstrating the potential benefits of online engagement for people in their real lives, which served as inspiration for later efforts to create automated health care intervention systems. The posts of online forum participants who self-disclosed their diagnosis of schizophrenia were studied by Haker et al. They also observed that afflicted users gained empathy and support, as well as information regarding drugs and contacting medical professionals, from other users. With the introduction of social media sites like Facebook, there are now more forums for conversation regarding mental health issues. As noted by Moreno et al., it can be difficult to diagnose cases of Major Depressive Disorder (MDD, henceforth referred to as "depression"), especially in older teenagers. Thus, from 2009 to 2012, they looked for Facebook accounts of first-year students whose status updates mentioned symptoms of depression. Following interaction with these students, a clinical screening was conducted to ascertain a diagnosis of depression, if the subjects were agreeable. More than twice as many students were at risk for depression if they had symptoms of depression in their status updates. Not only that, but the students' status posts mentioning symptoms of depression were frequently discovered to be a way to get attention or support, even while they were reluctant to get in-person help. Facebook depression disclosures were thus acknowledged as a valuable tool for identifying people who may have unmet needs for mental health care. This gave clear impetus to advance the techniques for identifying this condition early on in its course. As a result of the substantial body of work in this field that was produced during the pandemic, it is appropriate to review. This study focuses on techniques for identifying linguistic elements found in user-posted social media texts. Determining cutting-edge techniques for identifying linguistic traits linked to mental diseases is one of the primary goals of our review. This involves organizing data sets that have "ground truth" (gold standard) classifications of mental health state, which can be used to improve these techniques. EHRs, clinical surveys, and self-disclosure statements pertaining to mental health diagnoses (e.g., "I was diagnosed with depression") are some sources from which ground truths might be gathered. The study then investigates how longitudinal studies reflect the temporal stochasticity of mental states integrated into these approaches. Study also note similar ethical and technological limitations that were addressed during the preparation of the examined studies. Ultimately, recommendations will be made regarding the future course of AI-based mental health research.

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

Material and Methology A systematic review of the literature is a comprehensive investigation of the current and prior academic literature that employs a transparent and reproducible methodology for searching and synthesizing information with a high degree of objectivity. By utilizing the "Systematic Literature Review", this study ascertained the intellectual structure of the existing body of literature. Systematic review analysis is a well-established scientific method for analyzing literature. Although there are several methodologies for conducting literature reviews, SLR analysis is a widely favoured qualitative technique and yields the most effective responses to research inquiries (Amir, 2017). 2.1 Literature Search The search scope was restricted to conference papers, book chapters, and published research articles indexed in the Scopus database. The three most significant elements in the title or abstract, "mental health," "mental disorder," "mental illness", "schizophrenia", "social media", "twitter", "Artificial Intelligence" and "Facebook" were incorporated into the parameters for the literature search. Consequently, 1410 items were visually assessed during the preliminary screening phase. The criteria for the inclusion and exclusion of the articles under review are shown in Figure 1. of the research paper. This study only included articles that satisfied a pre-established set of criteria. For the bibliometric analysis, only food security, humanities, and food supply chain articles in an Indian context were selected. Therefore, only research articles published exclusively between January 1999 and January 2024 were included in this study. Google Scholar was used to look for published, peer-reviewed articles, which improved the literature search (Andalibi,2017). Finding research on the three primary mental health burdens—severe depression, anxiety, and schizophrenia—was the major goal of the search. A manual selection process was then used to assign each article to one of four subject areas covering significant and unique aspects of mental health research on social media: Data sets on Mental Health and Social Media, Methods for Predicting Mental Health Status, Longitudinal Analyses on Mental Health, and Ethical Aspects on the Data and Analysis of Mental Health. The Introduction's description of the objectives of this review is supported by these topics, which are covered in greater detail in Textbox 1.

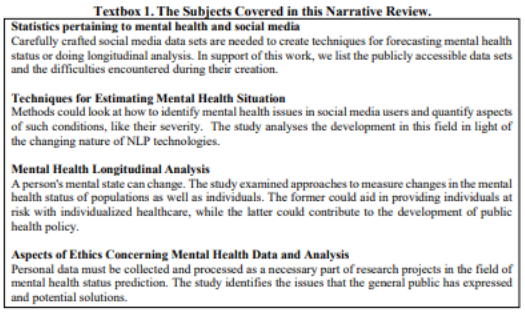

Figure 1. The subjects covered in this narrative review

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

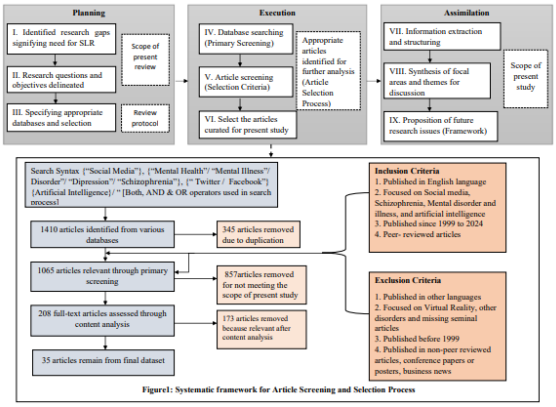

Figure 2. Systematic framework for article screening and selection process

II. METHODS

Approaches for predicting the mental health condition of social media users are evaluated in this study. For the analysis of vast quantities of written material, the study primarily concerned with the advancement of AI-powered techniques, specifically Natural Language Processing. Limitations that impact research in this field are also elaborated upon. The challenges encompass a scarcity of publicly available data sets of superior quality that may be utilised for methodological benchmarking, as well as the necessity to establish ethical and privacy frameworks that recognise the stigma and susceptibility of individuals impacted by mental illness (Yates, 2017).

Peer-reviewed articles published from 1999 to 2024 were retrieved via a Google Scholar quest. Data sets pertaining to social media and mental health, methodologies for forecasting mental health status, longitudinal analyses concerning mental health, and ethical considerations surrounding data and analysis of mental health were the four main areas of interest that we categorised the articles manually. The narrative evaluation comprised articles chosen from these groups.

III. FINDINGS AND DISCUSSION

The development of algorithms for forecasting mental health status, especially in the case of severe diseases like schizophrenia, requires larger data sets with the exact dates of participants' diagnosis. To address prevalent ethical and privacy apprehensions, researchers may benefit from soliciting users to contribute their social media data. Whatever the case may be, it seems highly probable that multimodal approaches to forecasting mental health status will yield progress that would be unattainable with artificial intelligence alone (Seabrook, 2018).

After a rigorous four-stage manual sorting process, 35 publications from all four subject areas were ultimately chosen for evaluation. The research conducted between 1999 and 2024 as well as significant occurrences like the pandemic were discussed in these publications. The articles that were ultimately selected for evaluation are detailed in Table 1. This table's layout was inspired by Szeto et al. Each of our four subject categories is covered in the following four sections, which give

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

a narrative review of these publications. Table 1. Articles from all four of the subject areas that were chosen for examination (Tadesse, 2019).

1. Social Media and Mental Health Data Sets

Access to good quality data sets is necessary to develop algorithms for mental health status prediction. Previous studies, according to De Choudhury et al., mostly focused on small, homogeneous samples of people who provided retrospective self-reports on their mental health, frequently through surveys (Resnik, 2015). It was acknowledged by the writers that an individual's social media postings could offer time-stamped insights about their psychological condition. In order to achieve this, they pooled the contributions of 476 Twitter users who self-reported as having depression to create a data collection of tweets. After that, the data was utilized to examine behavioral and language patterns, including diurnal activity and symptom mentions, respectively. The Center for Epidemiologic Studies Depression Scale screening test was one of the questionnaires that required manual completion in order to acquire the self-reported diagnosis, even though Coppersmith et al. regarded the data to be of excellent quality. In order to obtain self-reports of mental disease diagnoses on Twitter, such as "I was diagnosed with depression," they consequently developed an automated technique for labeled data set generation. A low cost and low resource strategy for data gathering was demonstrated by their yield of over 5000 distinct users who communicated such claims between 2008 and 2013. The writers did admit certain restrictions, though (Szeto, 2022). First of all, the sample consisted solely of Twitter users, which makes it unlikely to be representative of the broader public but comparable to other social media data sets nonetheless. Second, there was no way to confirm that the self-reported diagnoses accurately represented the psychopathology or were consistent with clinical diagnosis. For instance, self-reported depression is less heritable than depression diagnosed by a doctor, meaning that fewer genetic factors account for its variance in the population. This is supported by population biobank data. Nevertheless, it appears that this methodology served as the basis for a number of extensively utilized and publicly accessible mental health data sets. Reddit users who have self-reportedly diagnosed with depression through remarks such as the one above are included in the Reddit Self-reported Depression Diagnosis (RSDD) data set, which was generated by Yates et al. RSDD, which was further populated with 107274 non-depressed users for control purposes, has gained popularity as a tool for developing depression prediction techniques (Reece, 2017). The creation of two sister data sets, Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions (SMHD) and Temporal Annotation of Self-Reported Mental Health Diagnoses (RSDD-Time), has also been aided by it. After realizing that most study had not looked at the temporal aspects of mental health, MacAvaney et al. came up with the former. They manually annotated 598 postings from RSDD, which included a depressed user's self-reported diagnosis statement, at random to indicate the date of the diagnosis. In order to demonstrate a connection between certain time periods prior to diagnosis and the attitude a user displays in their postings, Owen et al. skillfully utilized RSDD and RSDDTime in a longitudinal study. Only 72 depressed users' posting histories

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

were used to base the findings, though, as many annotations in RSDD-Time indicate that it is impossible to estimate the diagnosis dates of many users with any degree of accuracy (for example, the user merely states that their depression diagnosis occurred "in the past") (Eichstaedt, 2018). This underscores the requirement for considerably larger data sets with highly accurate diagnosis dates for depression. On the other hand, SMHD emerged from the need for data sets encompassing a wide variety of mental health conditions (Thorn, 2023). Methods pertaining to depression, suicidal ideation, schizophrenia, and even multi-class experimental setups combining eating disorders, anxiety, ADHD, bipolar disorder, and PTSD were developed on this platform. Additionally, it was planned to gather a larger number of diagnosed users by utilizing a broader variety of higher positive predictive value patterns. Based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), these patterns identify diagnosis keywords pertinent to each disease. Consequently, 20406 diagnosed users and 335952 matched controls are present in SMHD. Despite these advantages, the lack of postings from mental health subreddits limits the utility of RSDD and SMHD. It is acknowledged that terminology used in subreddits devoted to mental health consistently varies from Reddit's general vocabulary. This must be continuously taken into account in research projects since it may bring bias into any models generated, along with the limitation that they only employed basic text patterns, like "I was diagnosed with depression," to gather users with mental health conditions. Data from social media also has additional biases. For instance, a greater proportion of men utilize the majority of social media sites, such as Instagram, Twitter, and Facebook. Additionally, research suggests that social media usage is more common among those with greater household incomes and educational attainment (Lan, 2019). Cohort-based data set design was examined by Amir et al. in order to reduce biases and enhance the representativeness of social media data sets. In other words, they created a pipeline for demographic inference that searched Twitter for users and determined their location, age, gender, and ethnicity in order to generate a subsample that was typical of the larger population (Dinu, 2021). Subsequently, they employed an established model to determine the frequency of depression and PTSD among the 48000 individuals gathered. This contrasts with the previously discussed method of potentially biased user identification—that of using self-reported diagnosis statement patterns. According to the authors, identifying population-level trends in the prevalence of disorders could be facilitated by the application of surveillance-based technologies. But they also noted that figuring out how social media data sets differ from representative samples of the general population will be necessary for a thorough assessment of these tendencies. Ethical and privacy concerns, in any event, limit the advancement and widespread use of surveillance-based techniques. For instance, obtaining consent from social media users would undoubtedly be necessary before automatically searching for and analyzing their data in large quantities, especially when it comes to personally identifiable information (age, gender, ethnicity, health status) and go into greater detail about these issues in the section on Ethical Aspects of Data and Analysis of Mental Health (Ammari, 2015).

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

2. Traditional Machine Learning Approaches

Methods for estimating mental health status from social media data started to surface in 2013, and these approaches frequently engaged interdisciplinary groups of clinical psychologists and computer scientists (Kessler, 2003). The use of supervised learning techniques to forecast depression in populations was advocated by De Choudhury et al. They created the social media depression index by utilizing post-level and user-level characteristics from a crowdsourced Twitter data collection. An SVM was employed in this process. Users' level of depression as expressed in their daily tweets could be ascertained with the help of the social media depression index. Women were shown to be 1.5 times more likely than men to display indicators of depression on social media in a US demographic population study. This finding slightly outperformed results from epidemiological surveys on formal diagnoses, which indicate that the figure is 1.3. It was suggested that techniques more attuned to language use could aid in the development of more robust models, since the overestimation was associated with women's higher emotional expressivity. Among these techniques is topic modeling using LDA (Kulkarni, 2023). Although the results of this approach, which has also been used to predict depression in Twitter users, should be interpreted with caution because the data set used in this study— which included both control and depressed users—was not considered to be representative of the general population. In subsequent research, depression and control users on Twitter were distinguished using LDA-derived features as input to an SVM classifier. Even though the topic-driven approach's usefulness was partially shown, just a meager 35 percent sensitivity result was obtained (Hruska, 2020). Using LIWC characteristics extracted from post text, Random Forests, another classic machine learning technique, was used in a comparable experimental setup for the prediction of depression in Twitter users. With the use of the Center for Epidemiologic Studies Depression Scale questionnaire, the 204 participants' mental health histories were gathered, yielding an impressive AUROC score of 87 percent (Ford, 2019). Emerging DL algorithms have the potential to improve techniques in this field and are expected to guide future research, according to Tsugawa et al. In the subsection that follows, go over these. A contemporaneous analysis also found that the idea of mass social media screening for at-risk persons was becoming closer in the near future due to advancements in NLP and ML (MacAvaney, 2018). In order to support this, it also highlighted two research that had a significant impact on data set design techniques and were covered in the previous section. The realization that these techniques could enable people to access the necessary medical care and social assistance earlier than they otherwise might has sparked interest in depression early prediction methods by 2019. Burdisso et al. created the SS3 algorithm, which determined the extent to which a given text fit into a particular category. This technique, which is broadly applicable to all domains, was employed in this instance to categorize and manage sad Reddit members. When compared to baselines generated using more conventional algorithms, like SVM, it showed better early risk categorization accuracy across multiple experimental scenarios. Additionally, it displayed computation times that were noticeably faster—roughly 20 times faster than SVM (Li, 2020). Giving its classification choices an explanation is one of SS3's other goals. It may help physicians by displaying relevant passages from

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

a user's Reddit content, like "Fact is, I was feeling incredibly miserable and wanting to kill myself." Traditional "black box" algorithms, such as SVM, are unable to provide this transparency. Because it does not always require processing the complete input text before delivering a classification judgment, SS3 was also praised as a low-resource technique, in contrast to SVM. It was acknowledged, nonetheless, that because it analyzes each word in the input text in a singleton fashion, it would not take into account potentially important two-word phrases like "murder myself" when making a classification choice (Liu, 2019).

3. Language Models and Transformers

NLP has a firm hold on language model capabilities by the early 2020s. Thus, building on the work of Burdisso et al., tweets indicating the presence of either depression or anxiety, or neither ailment, were used to apply BERT and ALBERT in an early depression prediction task. Because BERT and ALBERT have to consider the context of each word they encounter in a classification task, they also have to consider n-word phrases, resolving a problem that Burdisso et al. In contrast to an SVM baseline of 65 percent, an F1 of 77 percent was achieved using BERT in an experimental scenario where control and depressed users were balanced. However, in an imbalanced data set, BERT yielded 74% as opposed to 75% for SVM, which is a better representation of real-world settings where both technologies could be applied. Malviya et al. conducted a similar experiment in which they classified individual posts in a Reddit data set as depressed or not using BERT and traditional baseline techniques. Again, in a balanced experimental setup, strong BERT performance was observed, corroborating the notion that further research is required before LMs may be applied to this prediction task in more realistic, unbalanced environments. Some recommendations for achieving balance are resampling and synthesizing instances. A review of DL approaches to mental health prediction, carried out following the two studies, noted the value of the current data sets that we have already highlighted, but also stressed the need for more research with much larger data sets. Recently, some methods have utilized generative AI, specifically GPT. The field's opportunities have increased with the advent of generative AI. As we've already discussed, using high-quality data is essential to building methods for predicting mental health status. The lack of such data has led to the emergence of data augmentation techniques. Data synthesised from pre-existing data is a slightly different approach. As part of an annual workshop activity, a participating team used ChatGPT to compile data and build models for depressive symptoms as represented in Reddit posts that meet the criteria for the BDI-II. Countless texts that were suitable were generated. In reaction to the BDI-II response "I am so sad or wretched that I can't take it," ChatGPT generated the text "I'm so overcome by melancholy that I can barely function anymore." When it came to linking these texts to relevant BDI-II responses, real data models fared better than their synthesized equivalents (Singh, 2021). It was claimed that the synthesised texts were incredibly complex and detailed, which would have caused confusion for the learning machines utilized in the subsequent categorization experiment. One of the LMs employed was MentalRoBERTa, which was trained using genuine Reddit data. Future research should employ ChatGPT with greater caution to produce

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

fewer descriptive phrases that are more semantically similar to the BDI- II answers, as proposed. An other use of a GPT was in the SetembroBR Twitter corpus's automatic trisection of people who were depressed and in control. The GPT assigned each tweet a high, medium, or low relevance to mental health. The labeled data set was then fed into a bag-ofwords classifier, and the accuracy of the classifier's predictions was compared to a baseline produced by an earlier study that used BERT. It was stated that while this method, which was notably low resource-intensive, increased the sensitivity of the baseline result by 5%, more sensitivity might be obtained with improved GPT prompting—possibly by using a more rigorous definition of depression. Thus, there is optimism that LLM-assisted GPTs will support mental health prediction in multiple ways. To fully realize that promise, computer scientists need to consider how to improve GPT prompting tactics in each scenario. This section concludes with a review of the literature's treatment of schizophrenia. The aim of a 2015 study was to identify individuals with this disease from controls using LDA on a Twitter data set. The use of ambiguous phrases like "think" or "believe" and flat affect , which results from the absence of emoticons, were important discoveries. These were shown to be common in the postings of individuals diagnosed with schizophrenia. The inability to verify users' self-statements of their schizophrenia diagnoses was a restriction of their data collection, which is problematic in this area of research because psychotic symptoms may make patients less likely to believe in their diagnosis. In any case, because they can encounter stigmatizing comments on social media, people with schizophrenia might be reluctant to share their diagnosis. Using a human-machine paired strategy, Birnbaum et al. tried to achieve better accurate identification in Twitter. A doctor and a graduate-level mental health clinician examined the veracity of self-reported assertions about schizophrenia. The model that was subsequently constructed using machine learning demonstrated 87 percent sensitivity in differentiating between people with schizophrenia and controls. Notwithstanding this, the authors stated that access to the user's electronic health information is necessary in order to accurately establish the diagnosis of a user who makes a self-disclosure statement (Muhammad, 2023).

4. Longitudinal Analyses on Mental Health

The research that has been discussed thus far tends to predict a person's mental state at a given point in time. Though not static, a person's mental health is dynamic (De Choudhury, 2013). Indeed, some have argued that inferences made from "snapshots" of mental health problems at the sample level could not yield reliable predictions of how these states will change over time at the individual level. Consequently, research has also examined the temporal patterns of mental health disorders and symptoms. Facebook status updates from US college students may contain information that suggests depression symptoms, according to a 2011 study. It was noted that there were deficiencies in the identification and management of depression, particularly in university students. As a result, Facebook, (Moreno, 2011) a popular social media platform among students, provided inventive methods for identifying college students who might be in risk. 200 students' Facebook status updates from the prior year were collected manually. After reviewing the posts, human annotators

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

determined whether a depressed symptom was present if the Diagnostic and Statistical Manual of Mental Disorders criteria were satisfied. It was shown by terms like "hopeless" and "given up" that 25% of the profiles exhibited at least one depressive symptom. According to this data, Facebook may help identify students who are atrisk, which could open the door for more in-depth longitudinal research. The purpose of the study conducted by Schwartz et al. was to ascertain the annual variations in depression among Facebook users. From over 28,000 users' status updates, they retrieved 1-to 3-word keywords, topics derived from LDA, and LIWC categories. According to the results of a regression analysis, users had far higher levels of depression in the winter than in the summer, which is in line with findings from the literature on psychiatry ((De Choudhury, 2013). A baseline model that used the average attitude across all users' status updates was over three times more accurate than the best model, which only managed a thirty percent accuracy. Loveys et alresearch, .'s in contrast, involved predicting mental health conditions across much shorter time periods—hours, to be exact (Mikal, 2016). Tweets from over 2500 individuals who self-diagnosed as having schizophrenia or anxiety were automatically classified as positive, neutral, or negative. We tracked each user's attitude changes (or lack thereof) during the course of three consecutive tweets sent out over a three-hour period (Chong, 2018). These observations were dubbed "micropatterns" by us. It has been demonstrated that individuals with schizophrenia display less emotional variety in their tweets than control users. This data might point to a lack of emotive expression, which is a known characteristic of schizophrenia. Users who experienced anxiety were less likely than controls to tweet continuously in a positive manner, which is consistent with clinical studies. However, insufficient information from the micropatterns to assess the severity of the mental health illnesses could be resolved by improving the automatic categorization process and accounting for language aspects other than sentiment (such as terms that could be associated with particular symptoms) (Coppersmith, 2014). Research has also been conducted on emotions and how they change during the course of an internet post sequence. Seabrook et al. investigated the possibility that "emotion dynamics" on Facebook and Twitter could act as early warning indicators for depression risk. The possibility of using emotion instability and variability as a stand-in for depression severity as assessed by the Patient Health Questionnaire-9 (PHQ-9) was looked at. It was hypothesized that self-reported depression severity would be positively linked with negative emotion word fluctuation and instability across status updates. We collected the depression severity ratings and status updates from 29 Facebook users and 49 Twitter users. MoodPrism would track the participants' emotional posts and the severity of their depression (as determined by the PHQ-9) during a one-year period. According to the research, there may be a protective effect for mental health from greater variety in negative emotion expression on Twitter, and there may be a clue about the prevalence of depressive symptoms among social media users from the instability of negative emotion expression on Facebook (Coppersmith, 2014). These findings were, however, constrained by the incapacity to manually review the users' tweets for privacy issues. The lack of a manual verification process rendered the results nearly non-reproducible. A different study examined the possibility of using Twitter emotion displays to predict depression in 2018. Eight

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

basic emotions were identified in 585 depressed users' tweets over a four-month period: anger, contempt, fear, happiness, grief, surprise, shame, and confusion (Golder, 2017). The average strength of each emotion was determined using the EMOTIVE ontology and then applied to a time series analysis of each user. This investigation helped construct ML-based classifiers to identify whether or not previously undiscovered Twitter users were depressed. The best-performing Random Forests classifier configuration, when employing temporal characteristics, scored 87 percent sensitivity; however, when using basic LIWC data, the sensitivity was just 71 percent. This suggests that tracking a person's emotional changes over time may help identify depression in users. Thorough examination of the language used in tweets, including terms with ambiguity (like "maybe") and terms that are chronologically connected, may be able to predict its presence as well as its intensity. The onset of the COVID-19 pandemic coincided with the growth of transformer-based LMs. It was no coincidence that LMs would be mentioned at that particular time, as interest in methods for monitoring depression at the population-level on social media grew. Tweets from people who identified as depressed between March 3rd and May 22nd, 2020, were collected for one study (Kelley, 2022). The goal was to develop a model that would show how the depression levels of different groups changed as COVID-19 spread. They used the BERT-like model XLNET [130] and a geographic aggregation of users in the data set to demonstrate how depression levels fluctuated in New York, California, Florida, and the US during the aforementioned dates. Depression levels were found to be similar in all four geographic areas during the epidemic, rising steadily following the US National Emergency declaration on March 13, declining slightly after April 23, and then sharply increasing after May 10. Compared to the other two states and the national average, Florida had a far lower overall depression score. This may be due to Florida's normally lower rate of depression than the national average, even in the absence of the epidemic. Because only Twitter users were taken into account, these results are limited and do not fully represent the population. Owen et al. used LMs in a different way to try and ascertain the time interval between a Reddit user's depression diagnosis and the posts that most accurately represented their illness. The RSDD and RSDD-Time data sets were intersected to yield 56 depressed users and 168 controls. All user postings were taken into account in progressively larger temporal bands by BERT and MentalBERT, a specialist LM, up to 24 weeks (or roughly six months) prior to the depressed users' diagnosis dates. When 12 weeks of postings were taken into account, the LMs obtained F1-scores of 0.726 and 0.715, respectively. This suggests that the most poignant language used by depressed users occurs in the final three months before to their eventual diagnosis. The fact that the specialized LM is trained on material related to mental health may be the cause of its inferior performance compared to its general equivalent (Cohan, 2018). Posts in subreddits and other like places are not covered by RSDD. The diagnosis dates were only estimations, as was mentioned in the section on data sets on social media and mental health's RSDD-Time debate, which limited the findings. Regardless, the idea was made that a multimodal categorization strategy may yield more reliable outcomes. For instance, a Reddit user's post upvotes or downvotes may also be a good indicator of their mental health. We wrap up this section by going over the topics that have been

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

mentioned in the literature regarding schizophrenia once more. Hswen et al. looked into the language used by people with schizophrenia on Twitter to see whether it might be used to gauge suicide risk (Cai, 2020). They looked at the frequency of tweets about suicide and paid close attention to when these tweets occurred. According to their hypothesis, Twitter users who self-identify as having schizophrenia would be much more likely than Twitter users from the general population to post tweets that contain terms related to suicide, reflecting the higher risk of suicide seen in people with schizophrenia in real-world situations (Chen, 2023). Over the course of 200 days, the tweets of 173 control users and 203 individuals with schizophrenia were gathered. Since the term "suicide" is often used in debates about suicide, it should come as no surprise that only tweets containing those words were targeted. Most importantly, every tweet's time of day was noted. According to a logistic regression model, users with schizophrenia were far more likely than control users to tweet about suicide (odds ratio 2.15, 95 percent CI 1.42-3.28). Taking into account the timing of tweets, there was a strong correlation observed between the frequency of conversations about suicide on Twitter and those regarding depression and anxiety, a trend that is also consistent with published data. However, a major problem was mentioned, similar to research previously discussed, which was the inability to confirm the diagnoses of the schizophrenic users (Bucur 2023).

5. **Ethical Aspects on the Data and Analysis of Mental Health**

In order to enhance mental health prediction, people's privacy should be considered when developing data sets, coming up with plans, and conducting longitudinal analysis (Birnbaum, 2019). Finding out what people thought about Twitter's usage for population health monitoring was the aim of a study conducted in 2016 by Mikal et al. Depression was the focus of their qualitative inquiry. A focus group consisting of both Twitter users and non-users was assembled; some had received prior diagnoses for depression (Hirschberg, 2015). The group was questioned regarding privacy expectations and the potential for health monitoring to be done by machines. As long as user names were kept secret, participants generally felt that using publicly available data for health monitoring projects was acceptable. It was also noted that methods relying on simple keyword searches may not be entirely accurate, and that the results they yield may be misleading. It would be stigmatizing, according to participants, to incorrectly diagnose depression in a user whose identity is made public. The study was only suggestive because the group comprised of just 26 Twitter users who met certain demographic requirements (predominantly male with an average age of 26.9 years). However, additional evidence of concerns regarding this type of stigmatization was found in Conway and O'Connor's parallel study. Nicholas et al. address related privacy concerns. They note that the introduction of the General Data Protection Regulation in Europe, along with well-known events such as the Facebook data scandal involving Cambridge Analytica, put data privacy in sharp perspective (Holmes, 2020). The concerns of users are diverse (Ganguly, 2022). Some research findings on fear may have implications for credit card applications, employment prospects, and stigma . Data that has been deidentified may be re-identified through materials supplied with research articles, which raises concerns. According to Mikal et al. and Vornholt and

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

De Choudhury, anonymity appears to be especially popular. Thus, it's considered imperative to obtain the user's express consent before processing their data. One method is to agree to the terms and conditions set forth by social media companies. That being said, as these may not be read and understood, this may not constitute informed consent (Johnsen, 2002). One solution to the issue is to ask users directly to submit their social media data for research. Another recommendation is to include a feature that allows users to select whether or not to have their data used when they post it. In this area of mental health research, there is also an issue with the nomenclature used (GBD, 2022). Chancellor and colleagues investigated the way in which people are referred to in the literature in attempt to predict mental health status using social media data. Fifty-five articles had recurrent topics. In technical portions, human beings are referred to as "samples" and "data," while in introductions, they are sometimes referred to as "individuals" and "people." It has been argued that this might jeopardize scientific rigor as well as the populations the research is meant to benefit. Uncertainties regarding the study design brought about by conflicting terminology may affect the repeatability of outcomes. Depersonalization and dehumanization may also occur. In line with the previously reported research findings, this could result in the stigmatization of individuals and groups (Fujita, 1991). It is recommended that in order to mitigate this, more human-centered methods like participatory design be considered while conducting interviews and field research. That being said, this runs counter to the challenges listed in the section on social media and mental health data sets, which demonstrate how almost hard it is to collect large data sets with these methods. Välimäki et al. evaluated the available data and came to the conclusion that there is a paucity of research on the risks and attitudes around social media treatments for schizophrenia . Still, there are hints that some doctors are concerned that the condition bearer may get anxious if professional moderation is not applied to online peer assistance. Because cognitive deficits in people with schizophrenia can hinder the learning of digital skills, clinicians' concerns are warranted. Discussion and findings Techniques for predicting mental health status from social media data are gaining popularity, especially those that use natural language processing (NLP). There has been a discernible surge in interest in remote mental state monitoring of individuals and populations following the COVID-19 pandemic. In fact, compared to the previous 20 years, the search strategy used for this review produced 917 and 903 articles more in 2020–2021. In the past, text features were fed into typical machine learning algorithms. However, more advanced methods have been employed recently, such as transformer-based learning machines (LMs) and LLMs. The research community has made an effort to supply social media data in order to aid in this work, and it has done so in a way that is increasingly considerate of the subjects' ethical and privacy concerns. According to our investigation, depression is the most often reported ailment in publicly accessible data sets. This study also emphasizes the need for much bigger samples with precise documentation of contextual data, such the diagnosis date, rather than merely the condition's existence. Such data would probably corroborate findings from long-term research, the majority of which have also examined depression, and provide an extra window for prediction prior to a definitive diagnosis being made. The subject claims that gathering such ground truth data through conventional, private

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

polls is invasive and time-consuming (Picardi, 2016). One potential tactic, as examined by Eichstaedt et al., is obtaining authorization to access EHRs in order to accompany individuals' social media posts. In fact, this verification method is critical to research on schizophrenia because diagnosis self-disclosure statements can be imprecise, even with their high sensitivity. In any scenario, more social media information must be acquired in order to sufficiently support NLP techniques. Reddit data sets, for instance, must regularly incorporate postings from mental health subreddits with content from other subreddits. This will ensure that learning strategies that rely on pre-trained learning models on such data are less vulnerable to biases that could reduce their effectiveness. In the near future, LLM-driven technologies like ChatGPT and its offspring will probably support methods. Two such innovations that they have already spearheaded are the tagging of examples from mental health data sets and the production of synthetic data. While psychiatry research acknowledges that LLMs present opportunities, it also recommends longterm testing to identify which cues are best suited for a certain task. Instruction finetuning is another recommendation for improving LLM performance (Chancellor, 2019). We found that the field of population-level and individual-level longitudinal studies has not done enough research on the emotion portrayed in social media posts. With time, finer-grained language characteristics could be used to predict depression severity more accurately. The most promising approaches will probably be those that complement natural language processing (NLP) with multimodal techniques that consider non-text information from social media activity, as this will likely produce more insightful results. While this might involve timeaware image analysis of user posts on Reddit and emoji , it might involve considering user geolocations and profile photos on Twitter (Nicholas, 2020). Multimodal approaches may help allay some of the concerns our review has brought up regarding privacy. The public is concerned about the use of outdated keyword searches in mental health status prediction methods due to the potential for uneven outcomes. It makes natural that receiving the incorrect diagnosis of anxiety, sadness, or schizophrenia would insult a social media user. Therefore, multimodal approaches are being pursued to enhance the capturing of human behavior in real-world scenarios. Improving processes by themselves won't be sufficient to win over the people. Above all, user consent needs to be sought, and any study's data gathering procedures need to be made clearer. Broadly asking participants for permission to use their social media data for study may become standard procedure; this could happen perhaps when they make a social media post. A widely debated notion in the specialist literature holds that extended or other abnormal social media use patterns may either cause or exacerbate certain mental health problems. Despite not being the study's primary emphasis, our analysis of the quantity of published articles suggests that this related topic may need a further examination.

6. Potential Clinical Applications

The possible clinical uses of AI on social media data are now discussed, with reference to the studies included in this review (Owen, 2023). They consist of: 1) assessing population-level data to guide the development of health care services and policy; 2) identifying and making interventions

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

and support available to individuals who may be at risk of mental health issues; and 3) keeping an eye on current patients to spot early relapse signals and take appropriate action. In research on mental health status prediction techniques, the third application field was not wellrepresented. To better understand patients' experiences and perceptions of health services, as well as to identify patterns of risky behaviors across specific demographics, AI and NLP can be used at the population level to navigate large volumes of data and inform clinical needs in a particular area (for example, young people accessing accounts linked to pro-anorexia or encouraging self-harm). In the COVID-19 epidemic, NLP was employed, as previously mentioned, to assess vast amounts of social media data and pinpoint the particular issues that persons with mental illnesses faced, such as health fears, loneliness, and suicidality. Policy formulation and the distribution of resources in the health services are two areas where this kind of data might be useful. An important feature of this study is its speed, especially when compared to more conventional research approaches. This is useful when judgments need to be made fast, like in an unstable situation like a public health emergency. AI can help individuals and organizations identify those who are mentally ill or at danger of developing mental health issues so that early intervention services can be provided. As stated in the section on ethical aspects of data and analysis of mental health, there are certain worries about consent, data usage, and privacy. It's interesting to note that although both youth and mental health professionals felt that social media companies should use AI to proactively identify users who are at risk of suicide or self-harm and direct them to helpful information and resources, youth felt more strongly that AI should be used to promote helpful content like psychoeducation. The technical difficulties of achieving this include how localized health care providers can use the personal data gathered by international platforms to enhance service. Notwithstanding these obstacles, social media has shown to be a helpful resource for locating pertinent subjects for studies. Examples of this include treating young people with eating disorders and those who have witnessed suicide, as well as utilizing Facebook data to uncover relapses in schizophrenia patients.

7. Limitations

As far as predicting mental health status is concerned, we have reviewed the literature in four main areas. In addition to being the topic of potential review articles of their own, these areas offer opportunities for deeper coverage. Increased coverage could also be pursued, which would encourage more research. For instance, we have mostly covered research on natural language processing, sometimes giving consideration to multimodal alternatives. For data from primarily image-based platforms, like Instagram, visual computing offers methods that are applicable. For that reason, computer vision specialists might have a lot to say on this. The selection and analysis of articles is somewhat subjective due to the narrative review format. Our solution to this was to employ a clear search and selection procedure that was modeled after popular features seen in systematic reviews. Additionally, the articles we reviewed were limited to those in which the study participants selfreported a diagnosis of schizophrenia, depression, or anxiety; in general, however, any information gleaned from a social media post ought to be regarded as a self-report." The

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

psychopathological literature has examined the shortcomings of this approach. Despite the fact that it ensures that the input accurately represents the experiences and beliefs of the social media user, it also presents a challenge to automatically compile sizable data sets containing information about mental health positions. As an illustration, there are probably going to be more false positive cases of almost any common diagnosis compared to a manually assembled and curated data collection, however there may also be false negatives or controls that actually have a mental health condition. An even greater vulnerability for automatically generated data sets is created in the situation of schizophrenia, where the illness itself may contribute to the unreliability of self-reports, as explained above.

IV. CONCLUSIONS

Particularly in recent years, there has been a lot of interest focused on the scientific subject of mental health status prediction. It seems that the epidemic era sparked an increase in interest. Before approaches for assessing mental health state may be deemed trustworthy enough for therapeutic applications, more research must be done on them. Concerns concerning text-only strategies, especially those that depend on keyword searches, have been voiced by the general public. As previously mentioned, a lot of people use image-based social media sites like Instagram. So, multimodal approaches will probably be required to help win over the public's trust. They must so generalize to social media data that is based on text, voice, image, and video. To lessen the burden on medical and mental health services, the endeavor is justifiable. Integration of automated approaches and conventional approaches for early health care intervention may actually be beneficial. It is imperative that the ethical considerations around the acquisition and use of user data from social media platforms be given careful thought as this job cannot be done in a vacuum. Users should be asked for their consent, either by giving them the option to donate their social media data or by giving them the option to share their data on a post-by-post basis for research.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Dr. Sugandh Arora and Dr. Kirti Dang Longani for their invaluable guidance and support in the preparation of this article. Your insightful feedback, expert advice, and encouragement were instrumental in shaping this work. I am sincerely thankful for the time and effort you invested in helping me develop my ideas and for the knowledge you shared with me throughout this process. Your mentorship has greatly enhanced my understanding and ability to produce quality research. Thank you for your dedication and for being an inspiring educator.

REFERENCES

Ammari, T., & Schoenebeck, S. (2015). Networked empowerment on Facebook groups for parents of children with special needs. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2805-2814.

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

Amir, S., Coppersmith, G., Carvalho, P., Silva, M. J., & Wallace, B. C. (2017). Quantifying mental health from social media with neural user embeddings. Machine Learning for Healthcare Conference, 306-321.

Andalibi, N., Ozturk, P., & Forte, A. (2017). Sensitive self-disclosures, responses, and social support on Instagram: The case of #depression. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 1485-1500.

Birnbaum, M. L., Ernala, S. K., Rizvi, A. F., Arenare, E., Van Meter, A. R., De Choudhury, M., & Kane, J. M. (2019). Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. NPJ Schizophrenia, 5(1), 17. https://doi.org/10.1038/s41537-019-0085-9.

Bucur, A. M., & Cosma, A. C. (2023). Automatic detection and classification of mental illnesses from general social media texts. Proceedings of the International Conference on Recent Advances in Natural Language Processing, 358-366.

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F. M., Breen, G., Byrne, E. M., Clarke, T. K., Forstner, A. J., Grabe, H. J., Hamilton, S. P., Levinson, D. F., Lewis, C. M., Lewis, G., Martin, N. G., Milaneschi, Y., Mors, O., Müller-Myhsok, B., Penninx, B. W. J. H., Perlis, R. H., ... Kendler, K. S. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nature Genetics, 52(4), 437-447. https://doi.org/10.1038/s41588-020-0594-5

Chancellor, S., Baumer, E. P., & De Choudhury, M. (2019). Who is the "human" in human-centered machine learning: The case of predicting mental health from social media. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-32.

Chen, Z., Yang, R., Fu, S., Zong, N., Liu, H., & Huang, M. (2023). Detecting Reddit users with depression using a hybrid neural network SBERT-CNN. Proceedings of the 11th International Conference on Healthcare Informatics, 193-199. https://doi.org/10.1109/ICHI2023.193199.

Cohan, A., Desmet, B., Yates, A., Soldaini, L., MacAvaney, S., & Goharian, N. (2018). SMHD: A large-scale resource for exploring online language usage for multiple mental health conditions. Proceedings of the 27th International Conference on Computational Linguistics, 1485-1497.

Cong, Q., Feng, Z., Li, F., Xiang, Y., Rao, G., & Tao, C. (2018). XA-BiLSTM: A deep learning approach for depression detection in imbalanced data. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1624-1627.

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 51-60.

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 51-60.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. Proceedings of the International AAAI Conference on Web and Social Media, 7(1), 128-137.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Social media as a measurement tool of depression in populations. Proceedings of the 5th Annual ACM Web Science Conference, 47-56.

Dinu, A., & Moldovan, A. C. (2021). Automatic detection and classification of mental illnesses from general social media texts. Proceedings of the International Conference on Recent Advances in Natural Language Processing, 358-366.

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. Proceedings of the National Academy of Sciences, 115(44), 11203-11208. https://doi.org/10.1073/pnas.1802331115

Ford, E., Curlewis, K., Wongkoblap, A., & Curcin, V. (2019). Public opinions on using social media content to identify users with depression and target mental health care advertising: Mixed methods survey. JMIR Mental Health, 6(11), e12942. https://doi.org/10.2196/12942

Doktoronline. (2024, February 29). Klikk.no. https://www.klikk.no/helse/doktoronline/

Fujita, F., Diener, E., & Sandvik, E. (1991). Gender differences in negative affect and well-being: The case for emotional intensity. Journal of Personality and Social Psychology, 61(3), 427-434. https://doi.org/10.1037//0022-3514.61.3.427.

Ganguly, C., Nayak, S., & Gupta, A. K. (2022). Mental health impact of COVID-19 and machine learning applications in combating mental disorders: A review. Artificial Intelligence, Machine Learning, and Mental Health in Pandemics, 1-51.

GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry, 9(2), 137-150. https://doi.org/10.1016/S2215-0366(21)00395-3.

Golder, S., Ahmed, S., Norman, G., & Booth, A. (2017). Attitudes toward the ethics of research using social media: A systematic review. Journal of Medical Internet Research, 19(6), e195. https://doi.org/10.2196/jmir.7082

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. Science, 349(6245), 261-266. https://doi.org/10.1126/science.aaa8685

Holmes, E. A., O'Connor, R. C., Perry, V. H., Tracey, I., Wessely, S., Arseneault, L., & Bullmore, E. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science. Lancet Psychiatry, 7(6), 547-560. https://doi.org/10.1016/S2215-0366(20)30168-1.

Hruska, J., & Maresova, P. (2020). Use of social media platforms among adults in the United States—Behavior on social media. Societies, 10(1), 27. https://doi.org/10.3390/soc10010027

MacAvaney, S., Desmet, B., Cohan, A., Soldaini, L., Yates, A., Zirikly, A., & Goharian, N. (2018). RSDD-Time: Temporal annotation of self-reported mental health diagnoses. Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, 168-173.

Johnsen, J. A., Rosenvinge, J. H., & Gammon, D. (2002). Online group interaction and mental health: An analysis of three online discussion forums. Scandinavian Journal of Psychology, 43(5), 445-449. https://doi.org/10.1111/1467-9450.00313

Kelley, S. W., & Gillan, C. M. (2022). Using language in social media posts to study the network dynamics of depression longitudinally. Nature Communications, 13(1), 870. https://doi.org/10.1038/s41467-022-28513-3

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, A. J., Walters, E. E., & Wang, P. S. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). JAMA, 289(23), 3095-3105. https://doi.org/10.1001/jama.289.23.3095

Kulkarni, H., MacAvaney, S., Goharian, N., & Frieder, O. (2023). Knowledge augmentation for early depression detection. Proceedings of the International Workshop on Health Intelligence, 175-

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

191.
Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite BERT for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942. https://doi.org/10.48550/arXiv.1909.11942
Li, A., Jiao, D., Liu, X., & Zhu, T. (2020). A comparison of the psycholinguistic styles of schizophrenia-related stigma and depression-related stigma on social media: Content analysis. Journal of Medical Internet Research, 22(4), e16470. https://doi.org/10.2196/16470
Liu, G., Wang, C., Peng, K., Huang, H., & Cheng, W. (2019). Socinf: Membership inference attacks on social media health data with machine learning. Transactions on Computational Social Systems, 6(5), 907-921.
Mikal, J. P., Hurst, S., & Conway, M. (2016). Ethical issues in using Twitter for population-level depression monitoring: A qualitative study. BMC Medical Ethics, 17(1), 22. https://doi.org/10.1186/s12910-016-0105-5
Moreno, M. A., Jelenchick, L. A., Egan, K. G., Cox, E., Young, H., Gannon, K. E., & Becker, T. (2011). Feeling bad on Facebook: Depression disclosures by college students on a social networking site. Depression and Anxiety, 28(6), 447-455. https://doi.org/10.1002/da.20805
Muhammad, K. A. (2023). Unveiling the emotional and psychological states of Instagram users: A deep learning approach to mental health analysis. Information Sciences Letters, 12(5).
Nicholas, J., Onie, S., & Larsen, M. E. (2020). Ethics and privacy in social media research for mental health. Current Psychiatry Reports, 22(12), 84. https://doi.org/10.1007/s11920-020-01205-9
Owen, D., Antypas, D., Hassoulas, A., Pardiñas, A. F., Espinosa-Anke, L., & Collados, J. C. (2023). Enabling early healthcare intervention by detecting depression in users of web-based forums using language models: Longitudinal analysis and evaluation. JMIR AI, 2, e41205. https://doi.org/10.2196/41205.
Picardi, A., Lega, I., Tarsitani, L., Caredda, M., Matteucci, G., Zerella, M. P., & Biondi, M. (2016). A randomized controlled trial of the effectiveness of a program for early detection and treatment of depression in primary care. Journal of Affective Disorders, 198, 96-101. https://doi.org/10.1016/j.jad.2016.03.025.
Reece, A. G., Reagan, A. J., Lix, K. L. M., Dodds, P. S., Danforth, C. M., & Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. Scientific Reports, 7(1), 13006. https://doi.org/10.1038/s41598-017-12961-9
Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V. A., & Boyd-Graber, J. (2015). Beyond LDA: Exploring supervised topic modeling for depression-related language in Twitter. Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 99-107.
Seabrook, E. M., Kern, M. L., Fulcher, B. D., & Rickard, N. S. (2018). Predicting depression from language-based emotion dynamics: Longitudinal analysis of Facebook and Twitter status updates. Journal of Medical Internet Research, 20(5), e168. https://doi.org/10.2196/jmir.9267
Singh, A., & Singh, J. (2021). Automation of detection of social network mental disorders–A review. IOP Conference Series: Materials Science and Engineering, 1022(1), 012008. https://doi.org/10.1088/1757-899X/1022/1/012008.
Szeto, M. D., Barber, C., Ranpariya, V. K., Anderson, J., Hatch, J., Ward, J., & Coolman, T. (2022). Emojis and emoticons in healthcare and dermatology communication: Narrative review. JMIR Dermatology, 5(3), e33851. https://doi.org/10.2196/33851.
Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in Reddit

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo

social media forum. Access, 7, 44883-44893.

Thorn, P., La Sala, L., Hetrick, S., Rice, S., Lamblin, M., & Robinson, J. (2023). Motivations and perceived harms and benefits of online communication about self-harm: An interview study with young people. Digital Health, 9, 20552076231176689. https://doi.org/10.1177/20552076231176689.

Yates, A., Cohan, A., & Goharian, N. (2017). Depression and self-harm risk assessment in online forums. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2968-2978.

Zhang, Y., Lyu, H., Liu, Y., Zhang, X., Wang, Y., & Luo, J. (2021). Monitoring depression trends on Twitter during the COVID-19 pandemic: Observational study. JMIR Infodemiology, 1(1), e26769. https://doi.org/10.2196/26769

ISSN 2722-0672 (online), https://pssh.umsida.ac.id. Published by Universitas Muhammadiyah Sidoarjo