A Review of LWIC and Machine Learning Approaches on Mental Health Diagnosis

Ruh Sağlığı Hastalıkları Tanısında LIWC ve Makine Öğrenimi Yaklaşımlarının İncelenmesi

Bahar SERT[1]

Selami Varol ÜLKER[2]

Derleme/Review
Başvuru/Received: 19.08.2023; Kabul/Accepted: 05.09.2023

ABSTRACT

Machine learning methods are becoming increasingly popular in data analysis. In the field of mental healthcare, these methods provide support to mental disorder diagnosis. Pennebaker developed a dictionary-based text analysis program, and it is also used in mental health diagnosis. In this study, ML and Linguistic Inquiry Word Count (LIWC) studies conducted in the field of mental disorder diagnosis were examined. Researchers aim to integrate LIWC with machine learning to conduct more comprehensive studies. The objective of this study is to examine how combining ML and LIWC methods can detect mental disorder with a focus on comparative research. For this purpose, publications related to ML and LIWC in Google Scholar, Web of Science, Scopus, EBSCO, PubMed were examined. Studies utilizing machine learning and LIWC methods in mental health diagnosis were reviewed to establish an overview of the literature. A table summarizing 15 articles on integrating machine learning and LIWC for mental disorder identification was compiled. Subsequently, the working principles of machine learning and LIWC were examined, and research conducted in the field of mental disorder diagnosis was reviewed. Further research particularly those integrating or comparing these two methods needed to better understand machine learning and LIWC in mental disorder detection.

Keywords: LIWC, machine learning, mental disorders, psychology, text analysis

ÖZ

Makine öğrenmesi yöntemleri veri analizi alanlarında giderek popülerlik kazanmaktadır. Bu yöntemler ruh sağlığı alanındaki tanı belirleme çalışmalarına da destek sağlamaktadır. İlk olarak, Pennebaker sözlük tabanlı bir metin analizi programı geliştirmiştir ve bu program ruh sağlığı teşhisinde de kullanılmaktadır. Bu çalışma kapsamında ruh sağlığı hastalıklar teşhisi alanında yapılmış olan makine öğrenmesi ve Linquistic Inquiry Word Count (LIWC) çalışmaları incelenmiştir. Günümüzde daha geniş araştırmalar yapabilmesi için LIWC ile makine öğrenimini birbirine entegre etmek amaçlanmaktadır. Bu çalışmanın amacı, makine öğrenmesi ve LIWC yöntemlerinin birbirine entegre edilmesinin ruh sağlığı hastalıklarının teşhisinde etkisinin araştırılmasıdır. Özellikle karşılaştırmalı araştırmalara odaklanılmıştır. Bu amaçla, makine öğrenmesi ve LIWC ile ilgili olan Google Scholar, SAGE journals, Web of Science, Scopus, EBSCO, PubMed kaynaklarındaki yayınlar incelenmiştir. Literatürdeki genel durumun ortaya konması amacıyla, ruh sağlığı hastalıkları tespitinde makine öğrenmesi ve LIWC yöntemlerinden yararlanan çalışmalar derlenmiştir. Son olarak makine öğrenimi ve LIWC'in çalışma prensipleri incelenip ruh sağlığı hastalıkları alanında yapılan araştırmalar ve bazı çalışmalar tablolaştırılmıştır. Bu çalışmanın, ruh sağlığı hastalıkları tespitinde makine öğrenimi ve Dilbilimsel Sorgulama Kelime Sayımını daha iyi anlamak için özellikle bu iki yöntemi entegre eden veya karşılaştıran daha fazla araştırmaya ihtiyaç olduğundan, araştırmacılara faydalı olabileceği umulmaktadır.
Anahtar Kelimeler: LIWC, makine öğrenimi, ruh sağlığı, psikoloji, metin analizi

[1] Üsküdar Üniversitesi, M.Sc Nörobilim (Tezli), bahaarsertt@gmail.com, **ORCID:** 0009-0003-3806-97912
[2] Üsküdar Üniversitesi, Psikoloji Bölümü, selamivarol.ulker@uskudar.edu.tr, **ORCID:** 0000-0002-6385-6418

1. Introduction

In recent years, machine learning and Linguistic Inquiry and Word Count (LIWC) software have emerged as powerful tools in the field of Psychology, offering revolutionary solutions for mental disorder diagnosis and making significant contributions to human life. The concept of "slips of the tongue" was introduced by Sigmund Freud in 1901, which demonstrated that small speech errors could potentially reveal a person's true feelings or underlying thoughts (Ülker, 2022). Freud emphasized the importance of studying vocabulary and analyzing the traces it uncovers in examining the unconscious (Bilik et al., 2021). One of the most important applications of Pennebaker's function-word analysis is the detection of mental disorder via written texts. LIWC is a widely used computerized text analysis software that allows researchers to examine the emotional, cognitive, structural and process components found in language by counting the frequency of words in written text or speech (Thompson et al., 2023). Recent research indicates that the combination of LIWC and machine learning is becoming more prevalent in the diagnosis of mental disorder. Studies have focused on analyzing extensive amounts of text to establish connections between everyday language use and personality traits, social interactions, and cognitive patterns. Pennebaker claimed to have stumbled upon the considerable potential of computer-assisted text analysis in the early 1990s with the development of this program (Pennebaker and Chung, 2007). When combined with machine learning techniques, LIWC can enhance the performance of prediction systems. For instance, in a study on personality prediction on Twitter, LIWC was used as a linguistic approach to analyze language features and improve the system's accuracy (Salsabila & Setiawan, 2021). LIWC has additionally been employed in forecasting online mental health assistance-seeking patterns among the Covid-19 outbreak (Liu et al., 2022). These studies highlight LIWC's effectiveness in enhancing the performance of machine learning models in various applications. Machine learning algorithms can analyze large amounts of text data for text analysis in mental disorder diagnosis and identify patterns and linguistic markers revealing mental health conditions such as depression, schizophrenia, suicidal tendencies, etc. (Birnbaum, 2020).

As social media use has increased in recent years, research in this field has become easier, which leads researchers worldwide to initiate studies combining these two inventions. In this article, we have discussed the rise of machine learning in various fields in recent years and the extensive use of LIWC in mental disorder detection. This review article has examined studies on text analysis and mental disorder detection, which demonstrate the ability of LIWC and the conjunction of machine learning to analyze a specific text and reveal various social, cognitive and emotional processes. Also, a detailed table, which lists 15 publications looking into the effects of combining LIWC and machine learning techniques on the detection of mental disorder, has been provided at the end of this review. This article aims to inform researchers and practitioners about the potential benefits of integrating these two inventions in the diagnosis of mental disorders. By doing so, it may be possible to achieve more comprehensive and precise results and to facilitate further studies related to this integration.

2. Computerized Text Analysis: A Brief History

LIWC is a computerized text analysis tool developed by Pennebaker, which estimates the percentage of words in a given text that fall into one or more of over 80 linguistic and psychological categories suggesting diverse social, cognitive and affective processes (Tausczik and Pennebaker, 2010). LIWC analyzes written language based on word usage according to psychologically meaningful categories (Aghazadeh et al., 2022). By examining the linguistic content of a text, LIWC can provide insights into various aspects of language use related to mental disorders. The development of LIWC dictionaries has undergone significant changes over time. However, as computational power has advanced, researchers have sought to combine practical algorithms and statistical models. This collaboration aims to create a harmonious relationship between domain knowledge and computational approaches. The recent release of LIWC-22 has gone through considerable revisions and updates in comparison to its earlier versions. This dictionary has been developed by classifying the entries into two main categories: "Basic" and "Expanded." The Basic Dictionary comprising the majority of the dimensions found in previous LIWC versions admits potential modifications. The Expanded Dictionary, however, incorporates extensively updated iterations of conventional LIWC categories, along with a variety of novel categories and variables (Boyd et al., 2021). In several studies conducted in the field of mental health, it has been discovered that linguistic styles utilized in social media content can serve as indicators of mental disorders (Ramirez-Esparza et al., 2014). In recent times with the advancement of technology, machine learning has also been utilized to enhance and expand dictionaries. Machine learning allows the creation of broader dictionaries and more accurate determinations.

3. Previous Studies on the Automatic Prediction of Mental Disorder from Text

LIWC has been used in studies to find linguistic indicators of mental disorders in therapy sessions, since changes in mental disorders have been found to correspond with changes in language cues. For example, individuals experiencing improvements in mental disorders often demonstrate an increase in the use of positive emotion words and a decrease in negative emotion words. Additionally, researchers have used LIWC to identify individuals with depression by analyzing their language patterns. This includes analyzing their tendency to use first-person singular pronouns, the presence of negative emotion words in their statements, as well as the frequent use of words associated with sadness and anxiety (Burkhardt et al., 2021). For instance, Stirman (2001) conducted a study to explore whether distinctive linguistic features could be detected in those poems authored by poets who ended their own lives. The objective of this research was to assess two suicide models by employing a text-analysis software. Therefore, around 300 poems written by both suicidal and non-suicidal poets from various time periods beginning from the 1800s to date were examined using the LIWC computer program. There are various methods for assessing sentiment in language, ranging from basic dictionary-based techniques that classify words as positive or negative and count their occurrences, to more advanced approaches that consider the specific words used and their arrangement within a sentence or text. (Pennebaker et al., 2007). In their recent research, the main objective of Morales and Levitan (2016) was to identify different characteristics in speech patterns. In this study, these researchers discovered a strong connection between the level of depression and the usage of words related to work and sleep. This discovery is particularly intriguing because the words used in these two categories are commonly associated with various symptoms of depression such as fatigue, loss of energy, insomnia and hypersomnia. Additionally, the results of Robinson et al. (2022) regression analysis also indicated that the LIWC domains were important factors in predicting depression. Furthermore, the emotional tone had a moderate however, significant impact on suicidal thoughts. Lastly, results were found to be significantly associated with depression and it did not show any correlation with suicidal ideation. Regarding the detection of other mental disorder, diagnosing schizophrenia through text analysis, is the most challenging endeavor. A particular study focused on detecting and foreseeing early recurrence in individuals who had been hospitalized for schizophrenia, as opposed to the typical emphasis on identifying users with schizophrenia on social media. The author's integrated aspects of language usage and psycholinguistic attributes into their model by analyzing word patterns and using a language model (Bae et al., 2021). Previous studies have already shown links between these attributes and emotions, behaviors, and mental health conditions. However, to capture the way language is structured on social media, Bae et al. (2021) also included features like readability, word repetition and word length in their model. The findings suggest that social media activity can detect signs of psychotic relapse in young people who have recently developed psychosis (Birnbaum, 2019). Another study found that by using only tweets and bio text, the accuracy in predicting depressive symptoms has been 91% and 83% respectively. They also think that by focusing on a specific user domain or using a wide vocabulary, it is possible to enhance the performance even further (Safa et al., 2022).

Researchers have increasingly relied on text analysis softwares after the Covid-19 crisis to gain insights into various aspects of the outbreak such as mental health, public opinion, research priorities and diagnostic factors. These tools have been highly beneficial in examining vast amounts of textual data and extracting significant insights to support decision-making and research attempts.

4. Language Use During the Covid-19 Pandemic

Text analysis methods, such as LIWC, have been used by researchers to study the effects of the Covid-19 pandemic on people's lives. During the Covid-19 pandemic, the increased use of social media resulted in an abundance of data, and this led researchers to develop broader dictionaries on this subject (Kaur et al., 2021). By analyzing language patterns, LIWC provides valuable information about mental health conditions, public opinion, and media portrayal during this period. Numerous studies have utilized LIWC to examine the use of language throughout the pandemic (Su et al., 2020). These studies have particularly focused on analyzing expressions of loneliness on social media platforms such as Twitter (Guntuku et al., 2019). Various studies have discovered linguistic indicators that are linked to feelings of loneliness, anger, depression, and anxiety. This has offered valuable understanding into how individuals have been experiencing their mental health in the face of the pandemic. The LIWC tool has also been applied to examine how Covid-19 vaccine information is portrayed in the media and to examine the emotional and cognitive facets of individuals' vaccine experiences throughout the pandemic (Monzani et al., 2021). The increased use of social media has made it easier to access the data needed on these issues, and the creation of open dictionaries with the help of machine learning has encouraged scientists to investigate the data (Binjie et al., 2020). According to He and Cao (2018), the objective of these studies is to construct a predictive model for depression using only text-based content from social media platforms. This investigation also seeks to incorporate a broader set of linguistic characteristics associated with depression and shed light on the correlation between language usage and depression. The study utilizes several machine learning techniques like ridge, linear regression, support vector regression, random forest regression, and gradient boosting regression. The objective is to evaluate how effectively cultural and suicide-related lexicons may predict depression. By contrasting the results from the SCLIWC dataset, which contains the extra features, with those from the entire dataset, the usefulness of these lexicons was assessed (Huang et al., 2019). The Chinese Suicide Dictionary, the Chinese Version of the Moral Foundations Dictionary, the Chinese Version of the Moral Motivation Dictionary and the Chinese Individualism/Collectivism Dictionary were some of the linguistic resources used in this study to extract lexical features (Xing, 2018). Afterward, depression prediction machine-learning models were created using these features. The results demonstrated a moderate correlation of 0.33 between the projected scores and the actual scores. Suicide has also grown in importance as a public health issue in modern culture. Surveys or patients' spontaneous expression of their feelings and experiences are typically used in traditional ways of diagnosing suicidal ideation, which is seen as insufficient, inactive, and untimely. Considering that people seem more open to expressing themselves on social media sites like Twitter and Reddit, researchers have chosen to study these platforms in an attempt to address this problem (Massell, 2022). For instance, to extract language elements associated with suicide, researchers employed three separate dictionaries: a data-driven dictionary, a Chinese suicide lexicon and the LIWC dictionary. As a result of the construction of depression prediction machine-learning models and weak classification models using these features, six different detection outcomes were obtained as a result. To get the final weighted findings, logistic regression was used. According to the evaluation of these models, the recommended detection strategy significantly outperformed the methods currently used for feature selection. Overall, the researchers created a more effective technique for diagnosing suicide based on linguistic characteristics (Huang et al., 2022).

5. Performing Machine Learning Analysis with Linguistic Inquiry and Word Count 2015

In recent years, there has been growing interest in the diagnosis of mental disorders utilizing LIWC and machine learning. With the benefit of the text analysis tool LIWC, linguistic characteristics connected to mental diseases can be extracted. Furthermore, machine learning algorithms can discover patterns and connections between these language traits and the existence of mental disorders, enabling the creation of prediction models (Tausczik, 2010). Thus, numerous studies have, so far, looked into the possibility of using LIWC and machine learning to identify mental disorders by analyzing social media data. Safa et al., (2022) used a method known as crowdsourcing to gather information from Twitter users who had been given a clinical depression diagnosis. The researchers looked at a range of behaviors, including language use and linguistic preferences. The statistical classifier that was created using this data can determine whether a person is at risk of developing depression and according to research, social media can offer useful clues for spotting the onset of depression (Kelley, 2022). These warning signs include a decline in social interaction, a significant rise in unfavorable feelings and increasing anxiety about relationships and medications. (Chowdhury et al., 2021). Furthermore, the prediction of mental disorders has also been done using LIWC and machine learning. A context-aware transformer-based neural network was employed by Teferra and Rose (2023) to make predictions about Generalized Anxiety Disorder using impromptu speech transcripts. The results showed that this method outperformed the baseline logistic regression model that was based on LIWC. These findings demonstrate the potential of LIWC and machine learning in predicting and detecting various mental disorders. Furthermore, research has utilized machine learning techniques along with LIWC to estimate individuals' tendencies to seek online psychological assistance during the Covid-19 outbreak. These studies highlight the potential of LIWC and machine learning in predicting mental health-related behaviors and monitoring mental disorder (Liu et al., 2022; Enevoldsen et al., 2022). Vize et al. (2018) made predictions about the writing style of narcissists, expecting them to use disagreeable language in an extraverted, open-minded and masculine way. However, their data only supported the hypothesis that narcissists use open-minded language while the other predictions did not hold (Grijalva et al., 2015). The researchers also aimed to assess whether the LIWC could effectively detect narcissism. However, their results indicated that the LIWC profiles did not demonstrate proficiency in capturing narcissistic traits. This implies that psychologists who are interested in using language to understand and analyze narcissistic personality traits will need to employ Machine learning techniques, like the Cutler-Kulis model to achieve their goals. This may also apply to the broader literature on language and personality. As a future direction, Cutler et al. (2021) suggest that including the potential of machine learning in LIWC models will be an even more powerful tool for building a wider vocabulary task.

LIWC can address this issue by utilizing a pre-determined dictionary created by humans instead of relying on a machine-learning model. This dictionary allows text to be converted into a condensed and meaningful vector representation. However, due to the knowledge limitations of the dictionary editors, certain words that could provide valuable information may be overlooked. This becomes more prominent when analyzing informal languages, such as dialects, slang, or cyber language. Using machine learning techniques to identify effective solutions for mental health issues among social media users can address the issue faced by LIWC, which is the exclusion of potentially valuable words in informal language (Islam et al., 2018). When considering all these efforts, it is argued that this mechanism should be improved through the advancing technology. The table below compiles papers on mental disorder diagnosis by integrating LIWC and machine learning algorithms. This table may shed light on future studies.

Table 1

Integrated studies of LIWC and Machine Learning on Mental Disorder Diagnosis

Author	Year	Title	Design	Result
Cheng et al.	2017	Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study	Within-group Design	This conjunction proves to be beneficial in analyzing linguistic indicators of suicide risk and emotional distress in Chinese social media. It is capable of pinpointing distinct characteristics that deviate from what has been previously observed in the English literature.
Gaston et al.	2018	Authorship Attribution via Evolutionary Hybridization of Sentiment Analysis, LIWC, and Topic Modeling Features	Within-group design	The integration of the techniques discussed in this research significantly enhances the effectiveness of Authorship Attribution systems. Upon the selection of relevant features, the fusion of these feature sets outlined in this study achieves a perfect accuracy rate of 100% in correctly identifying authors.

Islam et al.	2018	Depression detection from social network data using machine learning techniques	Within-group Design	The author suggested that utilizing machine learning techniques, specifically Support Vector Machine and LIWC, can be a useful and easily implementable approach. The outcomes of these classifiers generally range from 60% to 80%.
Fatima et al.	2019	Prediction of postpartum depression using machine learning techniques from social media text	Within-group Design	The study found that the techniques used have a high ability to predict PPD content. The multilayer perception achieved 91.7% accuracy for identifying depressive content and up to 86.9% accuracy for predicting PPD content in holdout validation, outperforming support vector machines and logistic regression.
Marengo et al.	2019	Exploring the association between problem drinking and language use on Facebook in young adults	Within-group Design	The predictive capabilities of LIWC and LDA features were tested and compared in relation to users' risk of problem drinking scores. The study found that LDA features outperformed LIWC features in terms of

				their ability to predict problem drinking levels.
Glauser et al.	2020	Identifying epilepsy psychiatric comorbidities with machine learning	Within-group design	Research indicates that the use of larger datasets may improve the identification of anxiety and bipolar disorders. Using machine-learning to analyze spoken language could be a good screening option when traditional methods aren't feasible.
Pestian et al.	2020	A Machine Learning Approach to Identifying Changes in Suicidal Language	Between-group Design	This method investigates the consistency of suicidal language by utilizing advanced computational techniques. The findings indicate that a patient's language remains relatively similar 30 days after initial recording, whereas their responses to common assessment tools may vary. This can be valuable in the development of strategies that identify a subject's characteristics based on data analysis.

Bae et al.	2021	Schizophrenia Detection Using Machine Learning Approach from Social Media Content	Within-group design	This research indicates that by utilizing machine learning techniques in conjunction with natural language processing, it is possible to comprehend the linguistic traits of schizophrenia and identify individuals who have schizophrenia or might be prone to it through their social media posts.
Cutler et al.	2021	Inferring Grandiose Narcissism from Text: LIWC Versus Machine Learning	Between-group Design	Results suggests that the machine learning model successfully retained personality-related details, whereas LIWC did not. As a result, it implies that caution should be exercised in social-personality studies that solely rely on LIWC.
Weintraub et al.	2021	Using machine learning analyses of speech to classify levels of expressed emotion in parents of youth with mood disorders	Between-group design	Using machine learning to analyze speech features has demonstrated potential as an effective way to classify parents as either having high or low expressed emotions (EE), displaying promising results.

Taawab	2022	Detecting Self-Esteem Level and Depressive Indication Due to Different Parenting Style Using Supervised Learning Techniques	Within-group Design	The LR algorithm, using the count-vectorizer and LIWC, has outperformed the other two algorithms in terms of scores. For both the depression indication and self-esteem datasets, the LR algorithm achieved an accuracy and recall score of 83.00 and 76.80, respectively. Compared to the GBC and SVM classifiers, the LR algorithm showed superior performance when utilizing LIWC.
Pan et al.	2023	Linguistic Analysis for Identifying Depression and Subsequent Suicidal Ideation on Weibo: Machine Learning Approaches	Between-group design	Research findings indicate that the identification of anxiety and bipolar disorders could be improved by utilizing larger sets of data. Machine-learning analysis of spoken language shows promise as a viable screening option, particularly in cases where traditional methods are impractical.
Bartal et al.	2023	Identifying women with post-delivery posttraumatic stress disorder using natural	Within-group Design	Results indicates that birth stories have potential value in

		language processing of personal childbirth narratives		developing affordable and non-invasive methods for detecting maternal mental health issues. Further investigation utilizing machine learning and LIWC techniques to anticipate initial indications of maternal psychiatric disorders is justified.
Lyu et al.	2023	Detecting depression of Chinese microblog users via text analysis: Combining Linguistic Inquiry Word Count (LIWC) with culture and suicide related lexicons	Within-group Design	The findings indicated that every dictionary played a role in making accurate predictions. The linear regression model performed the best with a Pearson correlation coefficient of 0.33 between predicted and self-reported values, an R-squared value of 0.10, and a split-half reliability of 0.75.

Zhang et al.	2023	Detecting Narcissism From Older Adults' Daily Language Use: A Machine Learning Approach	Within-group Design	The findings demonstrated that the random forest classifier model attained a classification accuracy suggesting a satisfactory level of performance in classifying. By utilizing the random forest classifier, the significance of individual linguistic features in distinguishing between high and low narcissism was measured.

When reviewing the findings summarized in Table 1, it has been observed that integrating machine learning techniques with the LIWC structure leads to more accurate results in diagnosing mental disorders (Gaston et al., 2018; Islam et al., 2018; Bae, Shim, and Lee, 2021). According to some research, the limitation of the lexicon in text analysis techniques constitutes a prominent issue (Cutler et al. 2021). In the context of this problem, it has been shown that open-dictionary application with the help of machine learning techniques increases the reliability rate of diagnosis (Pestian et al., 2020). The studies in Table 1 show the distinction between LIWC and machine learning integrated studies in mental disorder diagnosis within-subject design as well as between-subject design. These studies also reveal that there was an excess of data that could be analyzed due to the increase in people's use of social media due to the Covid-19 effect (Lyu et al., 2023). The idea of obtaining a larger dictionary in order to draw meaningful conclusions from this abundance of data became widespread during this period (Bartal et al., 2023). In particular, machine learning algorithms such as the LR algorithm outperformed the other two algorithms in terms of scores when used with LIWC. For both the depression indicator and self-esteem datasets, the LR algorithm achieved an accuracy and recall score of 83.00 and 76.80 respectively (Taawab, 2022). According to the existing studies on the subject, elements affecting I, we, social, family, good and negative emotions, sadness, health, work, achieve, money and death significantly distinguished (Fatima et al., 2019; Lyu et al.,

2023) between those with depression and the control group with these variables accounting for 64% of the difference (Pan et al., 2023). In addition, LIWC is insufficient to diagnose complex disorders such as schizophrenia alone because the program has a human-made dictionary and it is not wide enough (Zhang et al., 2023). Bae, Shim, and Lee (2021) suggests that machine learning approaches combined with natural language processing could help understand the linguistic characteristics of schizophrenia and identify individuals with schizophrenia or otherwise at-risk individuals using social media texts. According to Table 1, this problem is caused by the fact that the machine learning model successfully preserves personality-related details, while LIWC cannot preserve these details (Cutler et al., 2020).

To sum up, machine learning algorithms and LIWC principles need to work in collaboration to reduce the margin of error in research. This way, the already abundant data can be processed quickly and reliably (Marengo et al., 2019). According to the articles in Table 1, there is a mass of data both due to the recent increase in the use of social media and because people express themselves in a more unfiltered manner on social media accounts. Thus, this field is very promising for further study and development (Islam et al., 2018).

6. Discussion

Text analysis is an effective method to detect behaviors. Moreover, they are particularly relevant to identify mental disorders. Recently, with the development of technology and the increasing use of social media, especially during the Covid-19 pandemic, researchers have recognized the necessity to improve text analysis systems and to integrate LIWC and machine learning techniques. This article has examined the studies conducted in this framework with particular attention to the trends and results of conjoining ML algorithms and LIWC, which are used by various researchers to address mental disorder detection. Indeed, dealing with the mental disorder detection, the conjunction of machine learning and LIWC seems to be an effective endeavor. Therefore, it is important to survey the integration of primary machine learning and LIWC models used in mental disorder detection.

Identifying mental disorders through text analysis involves several crucial elements. One fundamental aspect is the utilization of open-vocabulary methods which enable a detailed examination of language exceeding conventional analyses based on predefined word categories (Eichstaedt et al., 2021). This approach facilitates the recognition of relationships and trends in language that might not be captured by pre-established word categories. Another crucial aspect is the examination of extensive text datasets. According to research, it is important to analyze vast amounts of text like those in social media communications to find language variances linked to mental problems. Furthermore, it is critical to identify particular linguistic traits connected to mental disorders. For instance, specific words or phrases may signal signs of mental instability or despair (Liu et al., 2022). Researchers can create algorithms or models that automatically recognize and classify mental disorders in text data by analyzing such language markers. In addition, it is crucial to take the usage of the language into account. Text analysis should consider the specific mental health issues being targeted as well as the nuances and intricacies of the language. Understanding such details might improve the precision of detection and classification. Different mental disorders may be associated with unique linguistic patterns. For instance, researchers have advanced our understanding of narcissistic personality traits by employing machine learning approaches. The Cutler-Kulis model has demonstrated potential in accurately predicting narcissistic levels. One intriguing discovery of this model is that in line with the expectations, narcissists tend to have a language profile that is both agreeable and open-minded. It would be beneficial to move away from simple word-counting methodologies towards more sophisticated machine-learning language models to acquire a greater knowledge of how narcissistic people use language. (Cutler et al., 2021). For making predictions about the future, the combination of LIWC and machine learning offers a number of advantages, too. Firstly, using data-intensive machine learning methods like LIWC allows making decisions based on facts in a variety of industries, including marketing, manufacturing, healthcare and education. Machine learning algorithms are highly useful for determining future outcomes since they automatically enhance their performance through experience (Jordan and Mitchell, 2015). More accurate and reliable predictions can be made by combining LIWC, which provides linguistic and psychological insights with machine learning techniques. Compared to conventional methods like LIWC, machine learning models have outperformed them in a number of prediction tasks. Lee et al. (2019), evaluated the benefits and predictive power of LIWC, machine learning and human evaluations in predicting relationship intentions in online dating profiles. The results showed that machine learning models surpassed LIWC in terms of accuracy and performed comparably to human evaluations. Similarly, Cutler et al. (2021) indicated that machine learning algorithms outperformed LIWC in predicting personality traits, age and gender from text. These results clearly demonstrate the benefits of using machine learning for prediction tasks. In addition, compared to lexicon-based techniques like LIWC, machine learning models offer more flexibility and adaptability. Lexicon-based techniques have drawn criticism for their inflexibility (Wang et al., 2022). However, machine learning algorithms are more flexible for making future predictions since they are able to acquire data and adapt to various settings (Biggiogera et al., 2021). A further advantage of machine learning models over other deep learning methods is their ability to cope with nonlinear issues and reduce overfitting (Wang et al., 2022). A more complex understanding of complicated phenomena is also made possible by the union of LIWC with machine learning. For instance, Sundararajan et al. (2022) explored variations in cognition and emotion across two groups with different religious beliefs by using machine-assisted analysis, including LIWC. Their findings have practically replicated the findings of manual coding and provided a more thorough picture of the diversity of emotion and cognition. This highlights the benefit of using machine learning to increase the scope and depth of investigation. In sum, the combination of LIWC and machine learning has various benefits for future forecasts, including the ability to make decisions based on evidence, increased accuracy in comparison to conventional approaches, flexibility to various circumstances, and the capacity to comprehend complicated phenomena. Researchers and practitioners can improve their prediction abilities and gain insightful data by combining the qualities of LIWC with machine learning.

Bibliography

Aghazadeh, Sanaz, Kris Hoang, and Bradley Pomeroy (2022). "Using LIWC to Analyze Participants' Psychological Processing in Accounting JDM Research", *AUDITING: A Journal of Practice & Theory*, 41(3): 1–20. DOI: 10.2308/AJPT-2020-060.

Bae, Yi Ji, Midan Shim, and Won Hee Lee (2021). "Schizophrenia Detection Using Machine Learning Approach from Social Media Content", *Sensors*, 21(17): 5924. DOI: 10.3390/s21175924.

Bartal, Alon, Kathleen M. Jagodnik, Sabrina J. Chan, Mrithula S. Babu and Sharon Dekel (2022). "Identifying Women with Post-Delivery Posttraumatic Stress Disorder using Natural Language Processing of Personal Childbirth Narratives", *medRxiv : the preprint server for health sciences*. DOI: 10.1101/2022.08.30.22279394.

Bartal, A., Kathleen M. Jagodnik, Sabrina J. Chan, Mirithula S. Babu, & Sharon Dekel (2023). "Identifying Women with Postdelivery Posttraumatic Stress Disorder Using Natural Language Processing Of Personal Childbirth Narratives", *American Journal of Obstetrics & Gynecology MFM*, 5(3), 100834. DOI: 10.1016/j.ajogmf.2022.100834.

Biggiogera, Jacopo., George Boateng, Peter Hilpert, Matthew Vowels, Guy Bodenmann, Mona Neysari, ... & Tobias Kowatsch (2021). "BERT meets LIWC: Exploring State-of-the-art Language Models for Predicting Communication Behavior in Couples' Conflict Interactions. In *Companion Publication of the 2021 International Conference on Multimodal Interaction,* 385-389. DOI: 10.1145/3461615.3485423

Bilik, M. Zuhal, Eylül Ceren Hekimoğlu and Faruk Gençöz (2021). "Traces of the Unconscious in Language", *Language and Psychoanalysis*, 10(1): 27-35. DOI: 10. 1-9. 10.7565/landp.v10i1.4390.

Binjie, Gu, Weili Xiong, Zhonghu Bai (2020). “Human Action Recognition Based on Supervised Class-specific Dictionary Learning with Deep Convolutional Neural Network Features”, *Computers, Materials & Continua*, 62(3): 243-262. DOI: 10.32604/cmc.2020.06898.

Birnbaum, M.L., S. K. Ernala, A. F. Rizvi, E. Arenare, A. R. Van Meter, M De Choudhury, J. M. Kane (2019). “Detecting Relapse in Youth with Psychotic Disorders Utilizing Patient-generated and Patient-contributed Digital Data from Facebook”, *npj Schizophrenia* 5(17). DOI: 10.1038/s41537-019-0085-9.

Birnbaum, M.L., Raquel Norel., Anna Van Meter, Asra, F. Ali, Elizabeth Arenare, Elif Eyigöz, Carla Agurto, Nicole Germano, John, M. Kane, Guillermo A. Cecchi (2020). “Identifying Signals Associated with Psychiatric Illness Utilizing Language and Images Posted to Facebook”, *npj Schizophrenia*, 6(38). DOI: 10.1038/s41537-020-00125-0.

Boyd, Ryan L. and H. Andrew Schwarts (2021). “Natural Language Analysis and the Psychology of Verbal Behavior: The Past, Present, and Future States of the Field”, *Journal of Language and Social Psychology*, 40(1): 21–41. DOI: 10.1177/0261927X20967028.

Burkhardt, Hannah A., George S. Alexopoulos, Michael D. Pullmann, Thomas D. Hull, Patricia A. Areán, and Trevor Cohen (2021). “Behavioral Activation and Depression Symptomatology: Longitudinal Assessment Of Linguistic Indicators In Text-based Therapy Sessions”, *Journal of Medical Internet Research*, 23(7). DOI: 10.2196/28244.

Cheng, Qijin., Tim Mh Li., Chi- Leung Kwok, Tinhshao Zhu, & Paul Sf Yip. (2017). “Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study”, *Journal of Medical Internet Research*, 19(7), e243. DOI: 10.2196/jmir.7276

Chowdhury, Ekram Ahmed, Guy Meno-Tetang, Hsueh Yuan Chang, Shengjia Wu, Hsien Wei Huang, Tanguy Jamier, Jayanth Chandran, and Dhaval K. Shah (2021). “Current Progress and limitations of AAV Mediated Delivery of Protein Therapeutic Genes and the Importance of Developing Quantitative Pharmacokinetic/Pharmacodynamic (PK/PD) Models”, *Advanced Drug Delivery Reviews*, 170: 214–237. DOI: 10.1016/j.addr.2021.01.017.

Cutler, Andrew D., Stephen W. Carden, Hannah L. Dorough, & Nicholas S. Holtzman. (2021). “Inferring Grandiose Narcissism from Text: LIWC versus Machine Learning”, *Journal of Language and Social Psychology*, *40*(2): 260-276. DOI: 10.1177/0261927X20936309.

Eichstaedt, Johannes C., Margaret L. Kern, David B. Yaden, H. A. Schwartz, S. Giorgi, Gregory Park, ... and Lyle H. Ungar (2021). “Closed-and Open-Vocabulary Approaches to Text Analysis: A Review, Quantitative Comparison, and Recommendations”, *Psychological Methods*, 26(4): 398. DOI: 10.1037/met0000349.

Enevoldsen, Kenneth C., Andreas A. Danielsen, Christopher Rohde, Oskar H. Jefsen, Kristoffer L. Nielbo, & Søren D. Østergaard. (2022). “Monitoring of COVID-19 Pandemic-related Psychopathology Using Machine Learning”, *Acta Neuropsychiatrica*, 34(3): 148-152. DOI: 10.1017/neu.2022.2.

Fatima, I., Abbasi, B. U. D., Khan, S., Al-Saeed, M., Ahmad, H. F., & Mumtaz, R. (2019). “Prediction of Postpartum Depression Using Machine Learning Techniques from Social Media Text”, *Expert Systems*, 36(4), e12409. DOI: 10.1111/exsy.12409.

Gaston, J., Narayanan, M., Dozier, G., Cothran, D. L., Arms-Chavez, C., Rossi, M., ... & Xu, J. (2018, November). "Authorship Attribution via Evolutionary Hybridization of Sentiment Analysis, LIWC, and Topic Modeling Features", In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, 933-940. DOI: 10.1109/SSCI.2018.8628647.

Glauser, T., Santel, D., DelBello, M., Faist, R., Toon, T., Clark, P., ... & Pestian, J. (2020). "Identifying Epilepsy Psychiatric Comorbidities with Machine Learning", *Acta Neurologica Scandinavica*, 141(5): 388-396. DOI: 10.1111/ane.13216.

Grijalva, E., Newman, D. A., Tay, L., Donnellan, M. B., Harms, P. D., Robins, R. W., & Yan, T. (2015). "Gender Differences in Narcissism: A Meta-analytic Review", *Psychological Bulletin*, 141(2): 261–310. DOI: 10.1037/a0038231.

Guntuku, S. C., Schneider, R., Pelullo, A., Young, J., Wong, V., Ungar, L., ... & Merchant, R. (2019). "Studying Expressions of Loneliness in Individuals Using Twitter: An Observational Study", *BMJ open*, 9(11), e030355. DOI: 10.1136/bmjopen-2019- 030355

He, Lang, Cui Cao (2018). "Automated Depression Analysis Using Convolutional Neural Networks from Speech", *Journal of Biomedical Informatics*, 83:103–111. DOI: 10.1016/j.jbi.2018.05.007.

Huang, Jiaji, Qiang Qiu, Kenneth Church (2019). "Hubless Nearest Neighbor Search for Bilingual Lexicon Induction", In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4072-4080. DOI: 10.18653/v1/P19-1399.

Huang, Yan-Jia, Yi-Tin Lin, Chen-Chung Liu, Lue-En Lee, Shu-Hui Hung, Jun-Kai Lo, and Li-Chen Fu (2022). "Assessing Schizophrenia Patients through Linguistic and Acoustic Features Using Deep Learning Techniques", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:947-956. DOI: 10.1109/TNSRE.2022.3163777.

Islam, Md Rafiqul, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Hua Wang, Anwaar Ulhaq (2018). "Depression Detection from Social Network Data Using Machine Learning Techniques", *Health Information Science and Systems*, 6(1), 8. DOI: 10.1007/s13755-018-0046-0.

Jordan, Michael I. and Tom M. Mitchell (2015). "Machine Learning: Trends, Perspectives, and Prospects". *Science,* 349(6245): 255–260. DOI: 10.1126/science.aaa8415.

Kaur, Harleen, Shafqat U. Ahsaan, Bhavya Alankar and Victor Chang (2021). "A Proposed Sentiment Analysis Deep Learning Algorithm for Analyzing COVID-19 Tweets", *Information Systems Frontiers*, 23(6):1417–1429. DOI: 10.1007/s10796-021-10135-7.

Kelley, Sean W., Caoimhe N. Mhaonaigh, Louise Burke. *et al*. (2022). "Machine Learning of Language Use on Twitter Reveals Weak and Non-specific Predictions", *npj Digital Medicine,* 5(35). DOI: 10.1038/s41746-022-00576-y.

Lee, Chris, Tess V. Zanden, Emiel Krahmer, Maria Mos, and Alexander Schouten (2019). "Automatic Identification Of Writers' Intentions: Comparing Different Methods For Predicting Relationship Goals In Online Dating Profile Texts", Proceedings of the 2019 EMNLP Workshop W-NUT: The 5th Workshop on Noisy User-Generated Text, 94-100, DOI: 10.18653/v1/d19-5512.

Liu, Yali and Louisa Buckingham (2022). “Language Choice and Academic Publishing: A Social-ecological Perspective on Languages other than English”, *Journal of Multilingual and Multicultural Development,* Advance online publication DOI: 10.1080/01434632.2022.2080834.

Lyu, Sihua, Ren Xiaopeng, Du Yihua, and Nan Zhao. (2023). “Detecting Depression of Chinese Microblog Users *Via* Text Analysis: Combining Linguistic Inquiry Word Count (LIWC) with Culture and Suicide Related Lexicons”, *Frontiers in Psychiatry*, 14:1121583, DOI: 10.3389/fpsyt.2023.1121583.

Marengo, D., D. Azucar, F. Giannotta, V. Basile, M. Settanni (2019). “Exploring the Association between Problem Drinking and Language Use on Facebook in Young Adults”, *Heliyon*, 5(10), e02523. DOI: 10.1016/j.heliyon.2019.e02523

Massell, Johannes, Roselind Lieb, Andrea Meyer, and Eric Mayor (2022). “Fluctuations of Psychological States on Twitter Before and During COVID-19”, *PloS ONE,* 17(12). DOI: 10.1371/journal.pone.0278018.

Monzani, Dario, Alessandra Gorini, Davide Mazzoni, and Gabriella Pravettoni (2021). “Brief report – ‘Every little thing gonna be all right’ (at least for me): Dispositional Optimists Display Higher Optimistic Bias for Infection During the Italian COVID-19 Outbreak”, *Personality and Individual Differences*, 168:110388. DOI: 10.1016/j.paid.2020.110388.

Morales Michelle Renee and Rivka Levitan, “Speech vs. text: A Comparative Analysis of Features for Depression Detection Systems”, *IEEE Spoken Language Technology Workshop (SLT),* San Diego, CA, USA, 2016,136-143, (2016). DOI: 10.1109/SLT.2016.7846256.

Pan, Wei, Xianbin Wang, Wenwei Zhou, Bowen Hang, Liwen Guo (2023). “Linguistic Analysis for Identifying Depression and Subsequent Suicidal Ideation on Weibo: Machine Learning Approaches”, *International Journal of Environmental Research and Public Health*, 20(3): 2688. DOI: 10.3390/ijerph20032688.

Pennebaker, James W. and Cindy K. Chung (2007). “Expressive Writing, Emotional Upheavals, and Health”, H. S. Friedman and R. C. Silver (eds.), *Foundations of Health Psychology,* 263–284). Oxford University Press.

Pestian, John, Daniel Santel, Michael Sorter, Ulya Bayram, Brian Connolly, Tracy Glauser, Melissa DelBello, Suzanne Tamang, Kevin Cohen (2020). “A Machine Learning Approach to Identifying Changes in Suicidal Language”, *Suicide and Life-Threatening Behavior*, 50(5): 939-947, DOI:10.1111/sltb.126.

Ramírez-Esparza, Nairan, Adrian García-Sierra, and Patricia K. Kuhl (2014). “Look Who's Talking: Speech Style and Social Context in Language Input to Infants are Linked to Concurrent and Future Speech Development”, *Developmental Science*, 17(6):880-891.

Robinson, Eric, Angelina R. Sutin, Michael Daly, and Andrew Jones (2022). “A Systematic Review and Meta-analysis of Longitudinal Cohort Studies Comparing Mental Health Before versus during the COVID-19 Pandemic in 2020”, *Journal of Affective Disorders, 296*:567–576. DOI: 10.1016/j.jad.2021.09.098.

Safa, Ramin, Peyman Bayat, and Lelia Moghtader (2022). “Automatic Detection of Depression Symptoms in Twitter Using Multimodal Analysis”, *The Journal of Supercomputing*, 78(4): 4709–4744. DOI: 10.1007/s11227-021-04040-8.

Salsabila, Ghina and Erwin Setiawan (2021). “Semantic Approach for Big Five Personality Prediction on Twitter”, *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(4): 680-687. DOI: 680-687. 10.29207/resti.v5i4.3197.

Shuping, Xing (2018). “Analysis of the Characteristics of Advertising English Language”, *8th International Conference on Social Network, Communication and Education (SNCE 2018*). DOI: 10.2991/snce-18.2018.252

Stirman, Shannon. W., & James, W. Pennebaker (2001). “Word Use in the Poetry Of Suicidal And Nonsuicidal Poets”, *Psychosomatic Medicine*, 63(4): 517-522. DOI: 10.1097/00006842-200107000-00001.

Su, Yue, Jia Xue, Xiaoqian Liu, Peijing Wu, Junxiang Chen, Chen Chen, Tianli Liu, Weigang Gong, & Tingshao Zhu,. (2020). Examining the Impact of COVID-19 Lockdown in Wuhan and Lombardy: A Psycholinguistic Analysis on Weibo and Twitter. *International journal of Environmental Research and Public Health,* 17(12): 4552. DOI: 10.3390/ijerph17124552.

Sundararajan, Rajeswari, Preetha Menon, & Balaji Jayakrishnan. (2022). “Future of Artificial Intelligence and Machine Learning in Marketing 4.0. In *Proceedings of the 7th International Conference on Big Data and Computing*, 82-87. DOI: 10.1145/3545801.3545813.

Taawab, Al, A., Rahman, M., Islam, Z., Mustari, N., Roy, S., & Alam, M. G. R. Detecting Self-Esteem Level And Depressive İndication Due To Different Parenting Style Using Supervised Learning Techniques. *2022 9th International Conference on Behavioural and Social Computing (BESC)*. pp. 1-6, (2022). DOI: 10.1109/BESC57393.2022.9995147.

Tausczik, Yla R. & James W. Pennebaker (2010). ‘The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods’, *Journal of Language and Social Psychology,* 29(1): 24-54.

Teferra, B. G., & Rose, J. (2023). “Predicting Generalized Anxiety Disorder from Impromptu Speech Transcripts Using Context-Aware Transformer-Based Neural Networks: Model Evaluation Study”, *JMIR Mental Health*, 10, e44325. DOI: 10.2196/44325.

Thompson, Andrew D., & Maria Hartwig. (2023). “The Language of High-stakes Truths and Lies: Linguistic Analysis of True and Deceptive Statements Made during Sexual Homicide Interrogations”, *Legal and Criminological Psychology, 28*(1): 34–44. DOI: 10.1111/lcrp.12214.

Ülker, Selami Varol (n.d.). “The Associations between Self-Deception, Depressive Mood, and Attachment Dimensions with Linguistic Inquiry and Word Count”, *International Journal of Social Science and Humanities Research,* 5(8): 29. DOI: 10.47191/ijsshr/v5-i8-29.

Vize, Colin. E., Donald R. Lynam, Katherine L. Collison, & Joshua D. Miller. (2018). “Differences among Dark Triad Components: A Meta-analytic Investigation”, *Personality Disorders, 9*(2): 101–111. DOI: 10.1037/per0000222.

Wang, Yi-Chia, Robert Kraut, and John M. Levine (2012). “To Stay or Leave? The Relationship of Emotional and Informational Support to Commitment in Online Health

Support Groups", *CSCW '12 Computer Supported Cooperative Work,* Seattle, WA, USA, February 11-15. DOI: 10.1145/2145204.2145329.

Weintraub, M. J., Posta, F., Arevian, A. C., & Miklowitz, D. J. (2021). "Using Machine Learning Analyses of Speech to Classify Levels of Expressed Emotion in Parents of Youth with Mood Disorders", *Journal of Psychiatric Research*, *136,* 39–46. DOI: 10.1016/j.jpsychires.2021.01.019.

Wang, Xinyi, Sebastian Ruder, and Graham Neubig (2022). "Expanding Pretrained Models to Thousands More Languages via Lexicon-based Adaptation", In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),*863–877. DOI: 10.18653/v1/2022.acl-long.61.

Zhang, Shiyang, Karen L. Fingerman, Kira S. Birditt (2023). "Detecting Narcissism from Older Adults' Daily Language Use: A Machine Learning Approach", *The journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *78*(9): 1493–1500. DOI: 10.1093/geronb/gbad061.