**Social Media: Relation with Depression and its Detection using bagging classifiers**

**Keywords:** Social Media, Social Networking Sites, Depression, Bagging Classifiers, Mental Health.

**Abstract**

This study aims to identify social media and its relation with depression and how social media affects the mental health of individuals. The general Pakistani public who have attended college and are well educated is the study's target population. This research is based on a quantitative technique. A modified questionnaire was used in accordance with the study's objectives. The data was collected using Google forms. Five-point likert scales were preferred for the data collection when convenience sampling was used. The five-point Likert scale served as the foundation for the survey. The ADANCO software was used to carry out the testing. These tests include the convergent validity, discriminant validity, and Cronbach's alpha reliability and validity tests. ADANCO has been used to measure the path coefficient, adjusted (R2), and coefficient of determination (R2). The confined areas for answers in Pakistan were the main focus of this investigation. The sample size for this study is small relative to the population because it was completed in a short time. The findings of this study show that social media can eventually lead to depression. In this study, the elements that affect mental health by excessive use of social media were examined. Numerous studies have been conducted on the detection of depression by the use of social media through bagging classifiers. We have collected data on the detection of depression through bagging classifiers and have added it to our literature review.

**Introduction:**

The concept behind social media is not revolutionary. However, it appears that both professionals and academic researchers are unsure of exactly what should be categorized under this term and how Social Media differs from the closely related but seemingly interchangeable ideas of Web 2.0 and User Generated Content (Kaplan and Hanelin, 2010). CT Carr and RA Hayes (2015) strained to narrate the meaning of social media by using references from other researchers. According to them, several emerging definitions of social media have been put up, both within the transmission field and in connection with other fields, including public relations, information science, and mass media. Social media is often defined as digital technologies that place a strong emphasis on user-generated content or communication. Social media are frequently described in terms of their sound properties, either indicating the directionality of communications or utilizing particular tools, like Facebook or Twitter, to illustrate patterns of engagement. Despite the existence of numerous definitions, there is still no official, succinct, and widely accepted definition of social media.

Social networking sites (SNSs) have been absorbed into modern civilization, which could have an impact on mental health. (Seabrook et al., 2016). Balaji and Murthy (2019) propounded that the development of the web, which promotes, consumer generated media empowerment culture, usability, and interaction for end users, was aided by the expansion of social media during the past ten years. According to some estimates, about 4 billion individuals use social networking sites like Facebook, Twitter, and Instagram worldwide. This usage has led mental health professionals to look into whether social media's enormous popularity contributes to depression (Nittle, 2021). The use of social media is a significant risk factor for depression, according to researchers who are trying to understand how common it is in today's society (Hartanto et al., 2021).

Like other nations around the world, Pakistan has seen an increase in social platform users over time. Pakistan had a 36.5 percent internet crash rate at the beginning of 2022, or 82.90 million internet users and 71.70 million active social media users. With 71.70 million users, YouTube dominance the record of social apps in Pakistan, followed by Facebook with 43.55 million users. 3.40 million users, or 1.5% of the population, use Twitter, which is considerably fewer than other social media platforms. The amount of Twitter engagement did noticeably increase in March as various political events played out, according to an analysis of the prior several weeks' data stated by Arif (2022).

Scott and C. Woods (2019) reported that researchers studying sleep and mental health have been observing changes in the media environment over the past few decades, looking at the results of common uses of electronic media, such as watching TV, playing video games, and using smartphones. At the moment, the enhanced connectivity provided by devices and platforms has permitted not just an increase in the amount of time spent engaging in digital media activities, but also significant changes in social norms and expectations surrounding accessibility and online interaction etiquette. Almost all cohorts have seen a surge in social media use throughout the COVID-19 period, although younger individuals have been especially affected (M. Haddad et. al. 2021). Some research has proved that using social media excessively can have severe effects on a person, starting with anxiety and ending with depression (Bashir and A. Bhatt, 2017). ML techniques benefit the prediction and diagnosis in the healthcare domain by generating information from unstructured medical data. The prediction outcomes help to identify high-risk medical conditions in patients for early treatments. In mental disorders, ML techniques help arbitrate the potential behavioral biomarkers.to assist healthcare specialists in predicting the contingencies of mental disorders and administering effective treatment outcomes (Aleem et.al, 2022).

**Related Work**

This study of the literature aims to cite and compile evidence for social media research on depression symptom identification. In an analysis, DA Baker and GP Algorta (2016) provided a review of quantitative analysis for the relationship between depression and SNSs. A total of five databases were used. In all the studies that were reviewed, there were 196 participants in total. The study's global scope was demonstrated by the fact that it was conducted in 14 different nations, including the Philippines, Turkey, Serbia, Australia, Greece, and Korea. The participants' ages ranged from 15 to 88 years old. This review's objective was to explore and critically evaluate recent quantitative studies on online social networking and depression to better understand this relationship and to identify potential advantages and disadvantages of this behavior. The evidence points to the likelihood that how people utilize online social networking and the interpretations they give to their interactions play a significant role in determining their risk of developing depression, or vice versa. In a study, publications in English and German from 6 different databases were examined between 2010 and 2020. They showed that recent developments in big data analytics suggest analyzing language use and linguistic trends in social media material to forecast the occurrence of mental illness in the present or future through their analysis (N. Owusu et al., 2021).

A study conducted by Elizabeth et al. (2019) mixed method survey on public opinions to deduct depression through social media users. A cross-sectional, mixed-methods, web-based survey was sent out to SM users in the England who were 16 years of age or older. The study had 184 respondents in all, and 114 (62.3%) of them had previously struggled with depression. The researchers recognized the potential advantages of detecting depression from SM content but did not think the privacy dangers exceed these advantages. Most of the contributor found the analysis to be intrusive and exposing, and they would not provide their agreement for it to be done on their data. The studies described following the same aim and motive as ours. On Diverse segments of training samples, several algorithms are applied by Bagging classifiers as its works as an ensemble algorithm. The results of all algorithms' forecasts are then pooled. An improvement on bagging classifiers, Random Forest (RF) chooses characteristic pieces from the provided dataset in arbitrary combinations (Konieczny & Idczak, 2016). It has been developed on concepts of pooling and bootstrapping. With the utilization of training samples, bootstrap samples are developed. After that distinctive classifiers are trained with these bootstrap samples. The bootstrap data set frequently avoids incorrect training items. Additionally, aggregated classifier performance frequently outperforms that of a single classifier (Duin, 2003). Table 1 describes the Zulfiker et al.(2021) evaluation of bagging classifiers using diverse techniques and accuracy, sensitivity, specificity, precision and AUC of bagging classifiers have been demonstrated. In the view of Zulfiker et al.(2021) the bagging shows high precision in depression detection

Table 1:Utilization of diverse techniques for the classifier performance (Zulfiker et al., 2021)

| Classifier name | Technique | Accuracy | Sensitivity | Specificity | Precision | AUC |
| --- | --- | --- | --- | --- | --- | --- |
| Bagging | Without Feature Selection | 89.26% | 93.24% | 82.98% | 89.61% | 0.96 |
| SelectKBest | 90.91% | 90.54% | 91.49% | 94.37% | 0.96 |
| mRMR | 90.08% | 90.54% | 89.36% | 93.06% | 0.96 |
| Boruta | 90.08% | 91.89% | 87.23% | 91.89% | 0.96 |

**Table 2: Summary of Related Studies**

| **References of the studies** | **Methods and approaches used** | **Age of participants in the studies** | **Conclusion** | **Years Covered**  |
| --- | --- | --- | --- | --- |
| DA Baker & GP Algorta | Qualitative analysis by using five databases | 18 to 88.  | People using social media and SNSs are more likely to have depression. | 2010 to 2020 |
| N. Owusu et al. | Qualitative and statistical analysis by using 6 databases  | 18 or older.  | Extensive use of social media has provided mental problems to its users. | 2010 to 2021 |
| Elizabeth, et al. | Quantitative analysis by using mixed methods survey | 16 or older. | The content of social media has depressive features. | Not mentioned |

**Methodology**

This study's methodology is quantitative. The aim of this study is to determine the part of social media to cause depression and evaluating the literature about the bagging classifier role to identify the depression among people . Because askers with similar characteristics but from different geographic locations have been included, the testing strategy is cross-sectional (Gratton, C. & Jones, I., 2004). This paper's intended audience was any person who uses social media and the internet. The age range of the people studied in this study is 10 to 37 or above, and it is representative of the entire city of Lahore. We gathered opinions from 119 respondents.

Non-probability sampling and convenience sampling were used in this investigation. Good illustrationis a sort of sampling in which a sample is taken from the nearest region of the population. This kind of sampling is perfect for use in pilot studies. There are two sections to gather the information from the audience through a questionnaire that is adapted. The first section includes the age (10-37 or above), Gender (Male and Female). The questions about the study's variable were the focus of the second section. The scales utilized in this instrument's section varry from 1 (strongly disagree) to 5(strongly agree) on a 5-point Likert scale

This study is based on first-hand information collected using an adapted questionnaire and respondents from Lahore's general population. Literature, electronic books, electronic publication , and electronic articles were employed to gather data from derived sources. The information was assemble using Google forms. 119 submissions in all were submitted and evaluated using ADANCO.

**Results**

**Reliability and Validity**

Factor loading identifies the factor that has the biggest impact on each variable. The acceptability index ranges from 0.4 to 07, and the most significant influence on the variable of the study is indicated by the factor loading value of 0.7 (Mirza et al., 2020). The negligible impact of the factors on the variable will be in the case of the nearest 0 value of factor loading.

Table 3: Factor Loading

| Indicator | Social Media | Depression |
| --- | --- | --- |
| SM 1 | 0.7733 |  |
| SM 2 | 0.7149 |  |
| SM 3 | 0.5599 |  |
| SM 4 | 0.7342 |  |
| SM 5 | 0.7290 |  |
| SM 6 | 0.7119 |  |
| SM 7 | 0.6873 |  |
| SM 8 | 0.5543 |  |
| SM 9 | 0.6754 |  |
| SM 10 | 0.7325 |  |
| Depression 1 |  | 0.8262 |
| Depression 2 |  | 0.7995 |
| Depression 3 |  | 0.6695 |

**Internal Consistency**

The measure to determine the authenticity of the inner consistency of the scale is Cronbach's alpha. Its optimum value is 0.7)(Solution, 2022). The study is reliable as the value of Cronbach’s alpha is above 0.7.

Table 4: Cronbach’s alpha

|  **Construct Cronbach’s alpha(α)**  |
| --- |
| **Social Media** | 0.8763 |
| **Depression** | 0.7499 |

**Convergent Validity (AVE)**

The consistency of the manufacture evaluated utilizing Average Variance Extracted(AVE). The value of AVE 0.5 or above is referred to as the threshold but in different circumstances, a 0.4 value can be sufficient (Sablik et al., 2012).

Table 5: Convergent Validity

|  **Construct AVE**  |
| --- |
| **Social Media** | 0.4773 |
| **Depression** | 0.5900 |

**Discriminant Validity: Fornell- Larcker Criterion**

The level of the total difference of latent model variables has been extensively examined using the Fornell & Larcker (1981) criterion. The convergent soundness of the measurement model can be evaluated concerning this criterion using the AVE and joint confidence (Alarcón & Sánchez, 2015)

Table 6: Fornell- Larcker Criterion

| **Construct Social Media Depression** |
| --- |
| **Social Media** | 0.4773 |  |
| **Depression** | 0.4766 | 0.5900 |
|  |  |  |

**Coefficient of determination(R2) and adjusted (R2)**

The R2 threshold value is from 0-1 and it determines how much of the dependent variable reports the self supporting construct. The value should fall between 0 and 1. Depending on the type of research (Frey, 2022)

When values of adjusted R2 and R2 coefficients are below 0.2 then it indicates that from the perspective of a realistic approach, the mutual influence on latent constructs is weak (Sep & Wassertheil, 2014).

Table 7 depicts the values of both Coefficients of determination (R2) and Adjusted (R2) are above 0.2 so it depicts that the influence on the latent variable is strong enough.

Table 7: Structural Model

| **Construct Coefficient of determination (R2) Adjusted (R2)** |
| --- |
| **Depression** | 0.4766 | 0.4721 |

**Path Coefficient:**

The path coefficient also referred to as the beta value (β), shows how two variables are related to one another. Its range is from -1 to +1. A value of -1 denotes a negative correlation, while several +1 indicate a positive correlation. A 0 value indicates that there is no relationship(Hair et al., 2014). In the view of table 8, the association between dependent and independent variables is positive and remarkable as the value of β is 0.6904

Table 8: Path Coefficient(β)

|  **Independent variable**  | **Dependent variable** |
| --- | --- |
| **Depression** |
| **Social Media** |  **0.6904** |

**Graphical Representation of the Model:**

****

Figure 1: Graphical Representation of model

**Conclusion:**

The research related to the social media relationship to depression in people isn’t so enormous that’s why this study focused to recognize the role of social media to enhance the effect of depression among its users of Lahore and examine the utilization of bagging classifier and its accuracy level to determine the depression with respect to literature. According to the researchers of todays who are interested to understand the association of social media and depression demonstrate the social media as a notable risk factor to cause depression (Hartanto et al., 2021). This study considered diverse factors of social media such as frequent internet usage, less face to face interaction, cyberbullying, trolling, self-objectification to describe its impact on its users. The value of path coefficient ($β$) 0.6904 describes that the social media has significant and positive impact on its users. The literature explored that the machine learning approach, bagging classifier can be a phenomenal method of depression identification that can be a great use for the researchers of present era.

**References**

1. Alarcón, D., & Sánchez, J. A. (2015). Assessing Convergent and Discriminant Validity in the ADHD-R IV Rating Scale : User-Written Commands for Average Variance Extracted (AVE), Composite Reliability (CR), and Heterotrait-Monotrait Ratio of Correlations (HTMT). *Spanish STATA Meeting 2015*, 1–39.
2. Aleem S. et.,al (2022). Machine Learning Algorithms For Depression: Diagnosis, Insights, and Research Directions. *MDPI, Electronics DOI:* <https://doi.org./10.3390/electronics11071111/>
3. Andreas M. Kaplan and Michael Haenelin, 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Kelly school of business, Elsiver,* DOI: <https://doi.org/10.1016/j.bushor.2009.09.003>
4. Caleb T. Carr & Rebecca A. Hayes, 2015. Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication,* DOI: <https://doi.org/10.1080/15456870.2015.972282>
5. Duin, M. S. and R. P. W. (2003). Bagging, Boosting and the Random Subspace Method for Linear Classifiers. *Materials Science Forum*, *440*–*441*, 77–84. https://doi.org/10.4028/www.scientific.net/msf.440-441.77
6. Elizabeth M. Seabrook, et al., November 2016. Social networking sites and depression. *JMIR publications.* DOI: [10.2196/mental.5842](https://doi.org/10.2196/mental.5842).
7. E. Ford et,al. (2021). Public Opinions on Using Social Media Content to Identify Users With Depression and Target Mental Health Care Advertising: Mixed Methods Survey. *JMIR Publications,* DOI: <https://mental.jmir.org/2019/11/e12942/>
8. Frey, B. B. (2022). Statistical Power Analysis for the Behavioral Sciences. In *The SAGE Encyclopedia of Research Design*. https://doi.org/10.4135/9781071812082.n600
9. Hair, J. F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, *26*(2), 106–121. https://doi.org/10.1108/EBR-10-2013-0128
10. Hartanto A, Quek FYX, Tng GYQ and Yong JC (2021). Does Social Media Use Increase Depressive Symptoms? A Reverse Causation Perspective, *Front. Psychiatry 12:641934.* DOI: <https://www.frontiersin.org/articles/10.3389/fpsyt.2021.641934/full>
11. Hilal Bashir and Shabbir A. Bhatt (2017). Effects of Social Media on Mental Health: A Review. *The International Journal of Indian Psychology,* DOI: /10.25215/0403.134/
12. H. Scott and Heather C Woods, (2019). Understanding Links Between Social Media Use, Sleep and Mental Health: Recent Progress and Current Challenges. *Current Sleep Medcine Reports,* DOI: <https://link.springer.com/content/pdf/10.1007/s40675-019-00148-9.pdf>
13. Konieczny, R., & Idczak, R. (2016). Mössbauer study of Fe-Re alloys prepared by mechanical alloying. *Hyperfine Interactions*, *237*(1), 1–8. https://doi.org/10.1007/s10751-016-1232-6
14. Mirza, S., Sandhu, K., & Ameen, A. (2020). *Enhancing Relationship between Job Performance and Intellectual Capital through Organizational Commitment : An Evidence from Higher Education Institutes*. *9*(3), 590–600.
15. M. Haddad et.al, 2021. The Impact of Social Media on College Mental Health During the COVID‑19 Pandemic: a Multinational Review of the Existing Literature. *Complex Medical Pyschitric issues*, DOI: <https://link.springer.com/content/pdf/10.1007/s11920-021-01288-y.pdf>
16. N. Owusu et al. (2021). Artificial intelligence applications in social media for depression screening: A systematic review protocol for content validity processes. *PLOS ONE,* DOI: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0259499>
17. N Nittle, July 29021. How does social media play a role in depression. *Very Well Mind,* DOI: <https://www.verywellmind.com/social-media-and-depression-5085354>
18. P. Balaji, S. Sreenivasa Murthy (August 2019). Web 2.0: An Evaluation of Social Media Networking Sites. *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* DOI: <https://www.researchgate.net/profile/Pitchandi-Balaji/publication/340925331_Web_20_An_Evaluation_of_Social_Media_Networking_Sites/links/5ea473a945851553faaed260/Web-20-An-Evaluation-of-Social-Media-Networking-Sites.pdf>
19. Sablik, M. J., Rios, S., Landgraf, F. J. G., Yonamine, T., De Campos, M. F., Kim, J. H., Semiatin, S. L., Lee, C. S., Babu, J., Dutta, A., ABNT, Asm, A. N., Publication, I., Huang, J. C., Barnes, J. E., Williams, J., Blue, C. A., Peter, B., Asaadi, E., … Foram, Q. (2012). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Acta Materialia*, *33*(10), 348–352. http://dx.doi.org/10.1016/j.actamat.2015.12.003%0Ahttps://inis.iaea.org/collection/NCLCollectionStore/\_Public/30/027/30027298.pdf?r=1&r=1%0Ahttp://dx.doi.org/10.1016/j.jmrt.2015.04.004
20. Solution, C. dissertation by statistics. (2022). *Complete dissertation by statistics solution*. https://www.statisticssolutions.com/cronbachs-alpha/
21. S. Arfi (2022). Social media and politics in Pakistan: A major influence on people's’ opinions. *Pakistan Today,* DOI: <https://www.pakistantoday.com.pk/2022/05/16/social-media-and-politics-in-pakistan/#:~:text=At%20the%20start%20of%202022,million%20active%20social%20media%20users>.
22. Zulfiker, S., Kabir, N., Amin, A., & Nazneen, T. (2021). Current Research in Behavioral Sciences An in-depth analysis of machine learning approaches to predict depression. *Current Research in Behavioral Sciences*, *2*(April), 100044. https://doi.org/10.1016/j.crbeha.2021.100044