A Data-Driven Approach to Depression and Quality of Life Assessment Using Machine Learning

Paritosh Sahu, Naman Kumar Sharma and T Balachander

March 27, 2024

A data-driven approach to depression and quality of life assessment using machine learning

Paritosh Sahu	**Naman Kumar Sharma**	**Dr. T.Balachander***
Student - Final year,	*Student - Final year,*	*Asst. Professor,*
Department of NWC (E&T),	*Department of NWC (E&T),*	*Department of NWC (E&T),*
SRMIST, Chennai	*SRMIST, Chennai*	*SRMIST, Chennai*

*** Corresponding author**

***Abstract* -** This research proposal introduces an innovative application that utilizes machine learning to assess an individual's depression level and quality of life in real-time. The application offers personalized insights into the user's mental health and utilizes a distributed ledger system (DLS) that ensures data protection through decentralization and tamper-proof features, safeguarding it from unauthorized access [1]. Additionally, the app utilizes models to provide tailored recommendations and interventions for enhancing mental health and overall well-being [2]. Users will benefit from personalized guidance, including lifestyle adjustments, mindfulness exercises, and access to professional support.

I. INTRODUCTION

According to the World Health Organization (WHO), depression is estimated to affect more than 264 million individuals worldwide, and nearly half of these cases remain unattended [19]. The main reason for this therapy gap encompasses a lack of awareness, inadequate assessment methods, and a prevailing societal stigma surrounding mental health issues [18]. Alarmingly, the most significant portion of those affected with this debilitating condition remains undiagnosed and untreated due to a lack of awareness or reluctance to seek help [13]. In the context of this pressing public health issue, our research proposal proposes an innovative solution that utilizes machine learning models to address the barriers to mental health assessment and support for individuals.

Our endeavor is structured into two key components. The first part of our research focuses on the development of machine learning models designed to analyze user metrics, enabling the generation of personalized assessments of an individual's quality of life [1]. By harnessing real-time data, we aim to empower users with insights into their mental well-being. This initial phase of our study is a crucial step towards easing depression and facilitating early intervention.

In the second part of our experiment, our objective is to construct a model capable of offering personalized recommendations to users, fostering tangible improvements in their quality of life [2]. These recommendations may encompass lifestyle adjustments, mindfulness exercises, and access to professional mental health assistance. This personalized approach aims to not only increase awareness but also inspire positive action for those who may be unknowingly grappling with depression and try to keep them engaged with the recommendations without them getting bored of it.

A fundamental component of our research is the incorporation of a distributed ledger system whose inherent features of decentralization and tamper-proof data storage are crucial for ensuring the security and confidentiality of user information [3].

Our project's goal is to harness the potential of machine learning and other advanced technologies to bridge this gap. By providing a user-friendly, software-based solution, we aim to empower individuals to manage their mental health. This software will not only enable users to assess their own mental well-being through personalized metrics but also offer immediate access to professional consultations and tailored content designed to guide them on a path to recovery.

Our research seeks to revolutionize mental health care by offering a secure, efficient, and user-centric means of assessing, analyzing, and enhancing their well-being. By combining the power of machine learning and blockchain technology, we aim to enable individuals to make informed choices about their mental health. In doing so, we hope to contribute to a world where depression is better understood, more readily identified, and where effective support is both accessible and personalized.

II. RELATED WORKS

In recent years, machine learning and health care have witnessed significant progress, opening new perspectives on mental health and quality of life. This literature review examines key research papers regarding the impact of depression and Quality of Life and their contributions.

For doing this research we have used papers and concepts that help us in understanding how depression of a person was measured and his quality of life was calculated based on various factors and with those including from the simple daily lifestyle of a person to his external factors such as diet, work pressure, and other factors.

The base paper that we are utilizing for our project analyzes the depression-quality of life relationship, similar to our base paper another paper that we have referenced provided insights and the recommendations for improving the quality of life of a person [4]. Another study we conducted aimed at improving the mobile apps for student mental health support by addressing a few limitations. Furthermore, we used research that improvises the applications' security and mental health. Lastly, we utilized some papers that calculated the quality of life of a person from various pulmonary diseases using the nutrition from NHANES dataset [3].

The base paper we referred to used machine learning techniques like SVM and plotted pruned dendrograms using Self-Organizing Maps (SOM) which helped us categorize depression on how a person was affected. This multi-pronged approach contributes to a nuanced understanding of how depression affects individuals and, extending, their quality of life [4].

Our primary reference point, the base paper, rigorously analyzes the depression-quality of life relationship. We also analyzed a few papers that used various NLP models that were used to personalize recommendations from models like BERT from Google, etc. Further research from our side made us understand its limitations and think of implementing NLP in our research too for personalized recommendations and bridge the understanding of people from Machine learning using Natural Language Processing [2].

III. METHODOLOGY

A. User Metrics Collection:

The basic metrics collected from the users for this project are based on the questionnaire type of input which makes it easier to assess his situation and understand his depression level. For this questionnaire, we have taken the standard questions of PHQ-9 (Patient Health Questionnaire) as shown in Figure 3.1. the Stanford University prepared standard set of 9 questions that the user has to rate the question based on his experience in his past 14 days so for example let’s take an example like Feeling tired or having little energy this has to be rated by the user in his past 14 days of experience.

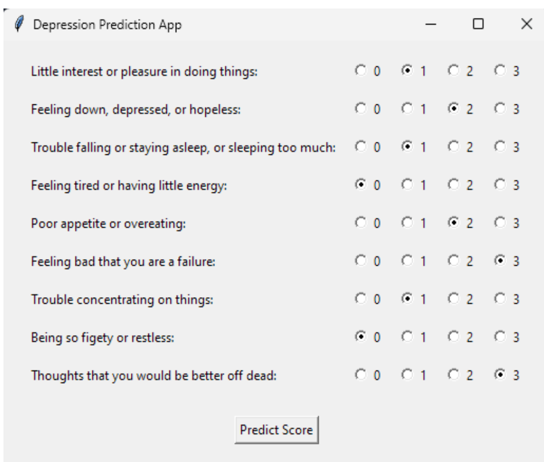


Figure 3.1 Initial parameters

Our initial score generation model dataset was made by collecting a survey that we passed on via forms which was later after giving importance values to each attribute which finally summed up generates our score Once the user inputs their questionnaire data, they will be prompted with their depression level which was categorized based on his score generated and be displayed to them. Then they will be prompted with an interface that shows them standard recommendations that a person must adhere to if his depression level is at the level the user scored from the questionnaire.

Later, our model suggests it, which is prompted to the user by asking him if they are satisfied with the original recommendation set or want it to be more personalized to them by first asking them what preset recommendations are helpful to them.

B. Data Analysis and Generation

Our recommendation model was also built up with the questionnaire model, which makes it easier from both sides to assess.

Person ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Scores
1	2	2	3	1	2	3	3	3	3	15.4
2	0	2	0	2	2	2	2	2	1	8.6
3	2	1	0	0	3	1	1	1	2	7.7
4	3	1	3	2	2	2	3	1	0	11.2
5	2	0	3	2	3	0	2	0	0	7
6	0	0	3	2	2	3	2	3	2	10.6
7	2	2	0	2	0	0	1	3	0	6.2
8	0	0	2	3	1	0	3	2	2	7.8
9	2	0	2	2	2	3	0	0	2	9.3
10	0	1	2	3	0	1	3	2	1	8.3
11	1	2	2	0	0	0	1	3	3	8.6
12	2	1	0	0	0	2	0	2	2	7
13	0	3	0	0	2	1	3	1	1	8.1
14	1	3	2	0	1	0	2	2	0	7.4
15	1	1	3	3	0	2	3	1	1	10.3
16	0	2	0	0	2	2	0	1	2	6.9
17	2	0	1	0	1	2	3	2	0	7.1
18	1	0	0	3	3	1	2	2	0	6.4

Figure 3.2 Dataset entities considered for asking Questionnaire

From figure 3.2, we can understand the important attributes from which a patient can be well understood on his mental health. Using the important attributes from each set of entities, we have prepared a set of personalized questionnaires to assess the user and build our recommendation set of parameters based on how the user responds to the questionnaire.

Our score generation model was trained on various classification models and amongst those decision trees with highest accuracy was taken to build up this model building.

This helped build up our model's accuracy and has made us use the decision tree model for the best accuracy. The recommendation model after the user answers his questionnaire, we have used the NLP model to properly categorize which recommendation must be given higher priority and what has affected our user the most and suggest recommendations.

Later we were also thinking of adding up the security aspects of the data too. If a new user tries to call upon his personal sensitive data, he might not be interested to share it with the general audience, so we thought of including the blockchain technology to further enhance the security of our model and thus completing the entire Analysis model.

C. Overall Workflow:

In this research, the steps taken to build on our model and how it works from the data collection step of data collection.

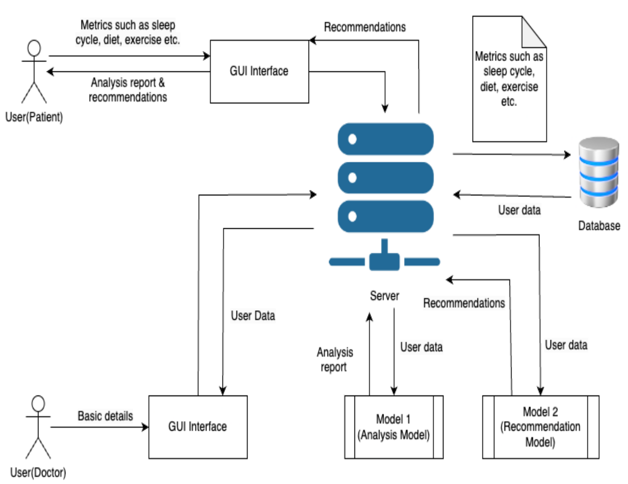


Figure 3.3 Architecture Diagram of the model

Initially, as shown in figure 3.3 in our model the user will be provided with an interactive application interface and is expected to provide the metrics about himself for the data models to analyze them and generate standard personalized recommendations for the user. So, our model was planned so that there are two separate users who use the applications, where one is the general user that wants to calculate their quality of life. Secondly, we have the Doctors or the medics as the other group of users having access to our application.

As the user enters their basic details for login into our application, the data they entered is stored in our databases securely using the servers. The servers that are being used to store and retrieve the data are connected with our analysis and recommendation models, where in which the analysis model understands the patients' conditions and the recommendation model after taking the report from the analysis model generates a set of recommendation for the user to make in his day-to-day life activities and suggest some mindful activities to keep the user engaged in positive activities and help improving their quality of life categorically.

Our recommendation model will be recommending to those who are experiencing the severity of their depression at a range of minimal, mild, moderate, moderately severe and Severe depression. The other side note if the condition of the user is severe and moderately severe, that is where the second user's role comes into play.

The recommendation model in these cases suggests the user to better get some expert advice on the state of his mental health. [13] As in these cases if the level of depression is severe and the user is having some illness like those of hallucinations and other which can only be treated via medications or needed to be supervised by doctor on his body conditions before suggesting any recommendations as the usual recommendations might not be effective on this kind of patients and make his suffer further than curing him.

At the last point, as the recommendations have been validated, the correct set of recommendations that are actually effective on the user are targeted based on his interests and are made engaging so that they don't start discontinuing it once more.

IV. RESULTS AND DISCUSSIONS

For understanding and classifying users after training the dataset, we must first use classification methods to classify the severity of a person's depression and provide a general set of recommendations accordingly. We used the decision tree model that belongs to a family of supervised learning. Using supervised learning, we used our model to train it to classify the level of depression a user suffers in 5 levels by using which we can give them recommendations.

We also tried using various classification methods, like the support vector machine (SVM's), random forest and other classification techniques. But the decision tree model we used provided higher accuracy to our prepared dataset, and we trained it well according to our needs for rating reference figure 4.1.

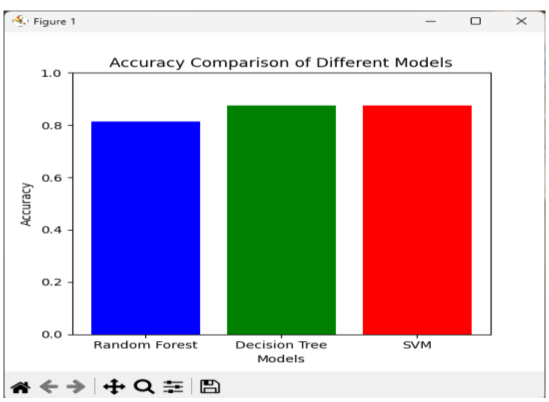


Figure 4.1 Accuracy comparison with other classification models

We created a dataset by collecting data from around 300 users trained in the decision-tree classification model. We trained our model with a training to testing ratio as 70 to 30 and predicted the incoming values the users will enter. The model utilized has been successful in terms of performance metrics, having an accuracy of approximately 83%, which can be concluded after a considerable number of iterations.

Since it is a classification model the optimum accuracy is only achieved after a certain number of iterations and it is noted that the highest percentage of accuracy for the depression calculation model at a certain point is 96.82%. The accuracy was calculated on the results of all the test data's outputs. To achieve that level, we need to train our model with a large amount of data only after which our accuracy in predicting the individual's true values could be determined.

As for the interface of the project as for now we have used the Tikinter which is a python library used to build a simple GUI interface it's also used for creating graphical user interfaces (GUIs) with various widgets such as buttons, labels, text boxes, etc. being easy to use.

The GUI has the fields for the values to which the user has to enter the data for and later in which the data after clicking the submit button at the bottom of the screen as shown in figure 3.1, it will display the user's depression rating and the level of depression user is suffering refer the below code in figure 4.2.

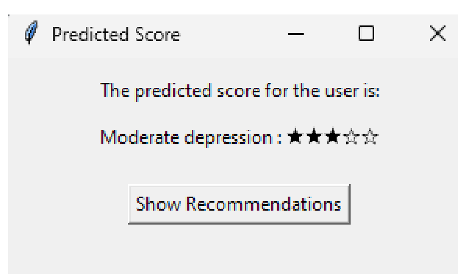


Figure 4.2 Feature Importance

After the score being generated the user will be shown an option to show recommendations which is personalized based on their score and a general set of recommendations are shown as per his depression score like for the result in Fig 4.3 the general set if recommendations that will be displayed to such user will be like, explore outdoor activities for a change of scenery, consider group therapy session. Here, the interesting aspect is that the user can choose which of the proposed recommendations is/are beneficial for him, so while personalizing his final recommendation we can include these activities similar to that he has highlighted useful.

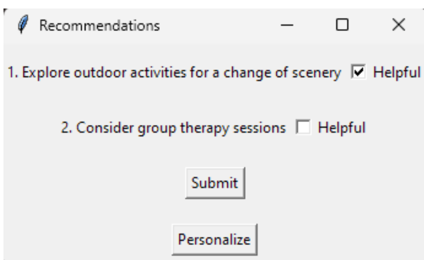


Figure 4.3 generalized recommendation

All the inputs the user selects are carefully noted and captured for personalization. Now if the user is interested to tailor his recommendations further, he can select the personalize button displayed to him as shown in figure 4.3 from where the use will be asked a general questionnaire from the important entities used to assess a patients mental health refer figure 3.2 for knowing what were considered important for making this personalizing questionnaire.

Once the user clicks on Personalized the user is first asked with few sets of questionnaires, they range from selecting few checkboxes on why they have planned to take the survey today and few questions asked in the form of either yes or no and few in the form of good, fair and bad format making it easy to answer refer figure 4.4.

As our personalizing questionnaire has been completed, we can proceed without final recommendations which are built upon the user answered metrics from the personalization questionnaire where we use the NLP for better understanding of the user's response and to understand what recommendations to provide. We have also iterated with several NLP models and found one which satisfies our needs.

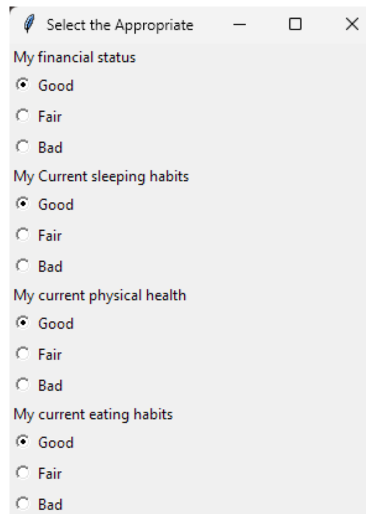


Figure 4.4 Personalizing recommendations

We used the NLP model of Vader lexicon from the nltk python library which is majorly used for sentiment analysis. The main reason for us going for this model is its faster response speeds than other NLP models side by side. Hence by tracking the user responses we do a sentiment analysis based on the score that is generated for each of his responses as shown in figure 4.5 and categorize them for it being either positive or a negative sentiment for a respective response from the user.

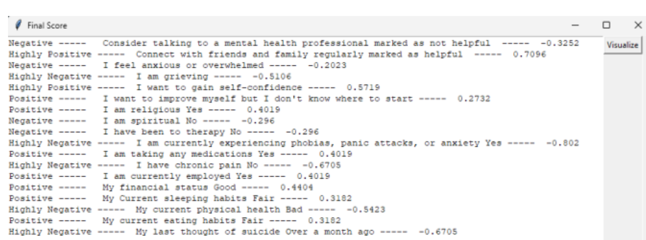


Figure 4.5 Sentiment analysis

From these responses, the sentiments being categorized, we take the polarity scores of the sentiments that are 5 most negative to visually show the user's parts for improvement.

The Visual representation is followed up to display it in a pie chart to make it easier to comprehend from the user as shown in figure 4.6 and better understand themselves.

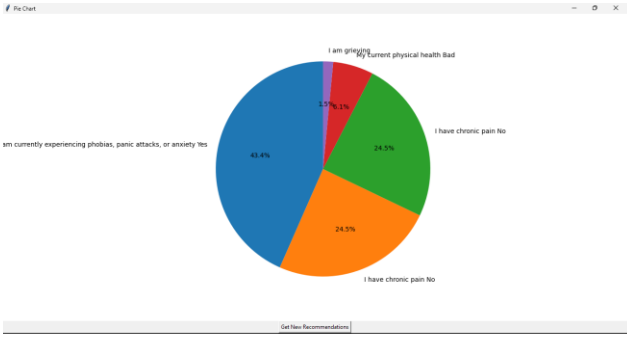


Figure 4.6 Visual Representation

Once this is done based on his polarity scores and the user's responses, we can obtain the new recommendations from a dataset briefly making him understand where he can improve his quality of life and make him engaged in the activities. Adding further to this at the end all the selections that the user had made were tracked was analyzed by our dataset model upon clicking the get new recommendation button a final summarized recommendation will be generated combining all the factors as shown below in figure 4.7 which was generated by t5-small NLP model which was pretrained for tokenizing and generating a summarized recommendation from the prepared dataset refer figure 4.9 for the dataset.

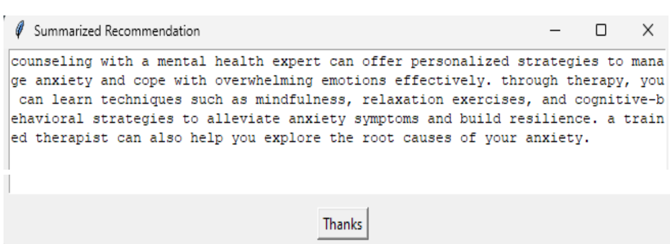


Figure 4.7 Final Recommendation

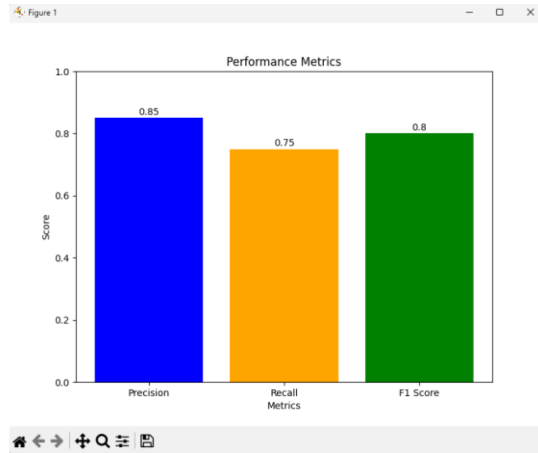
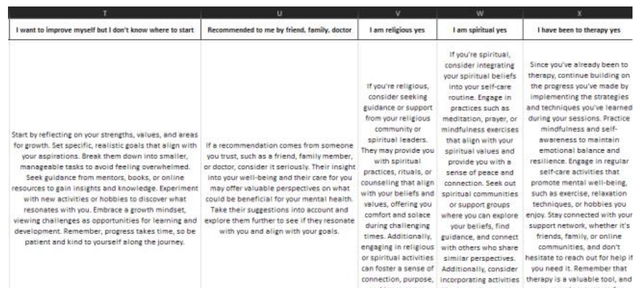


Figure 4.8 Precision, Recall and F1 scores



T	U	V	W	X
I want to improve myself but I don't know where to start	Recommended to me by friend, family, doctor	I am religious yes	I am spiritual yes	I have been to therapy yes
Start by reflecting on your strengths, values, and areas for growth. Set specific, realistic goals that align with your aspirations. Break them down into smaller, manageable tasks to avoid feeling overwhelmed. Seek guidance from mentors, books, or online resources to gain insights and knowledge. Experiment with new activities or hobbies to discover what resonates with you. Embrace a growth mindset, viewing challenges as opportunities for learning and development. Remember, progress takes time, so be patient and kind to yourself along the journey.	If a recommendation comes from someone you trust, such as a friend, family member, or doctor, consider it seriously. Their insight into your well-being and their care for you may offer valuable perspectives on what could be beneficial for your mental health. Take their suggestions into account and explore them further to see if they resonate with you and align with your goals.	If you're religious, consider seeking guidance or support from your religious community or spiritual leaders. They may provide you with spiritual practices, rituals, or counseling that align with your beliefs and values, offering you comfort and solace during challenging times. Additionally, engaging in religious or spiritual activities can foster a sense of connection, purpose, and inner peace.	If you're spiritual, consider integrating your spiritual beliefs into your self-care routine. Engage in practices such as meditation, prayer, or mindfulness exercises that align with your spiritual values and provide you with a sense of peace and connection. Seek out spiritual communities or support groups where you can explore your beliefs, find guidance, and connect with others who share similar perspectives. Additionally, consider incorporating activities like journaling, nature	Since you've already been to therapy, continue building on the progress you've made by implementing the strategies and techniques you've learned during your sessions. Practice mindfulness and self-awareness to maintain emotional balance and resilience. Engage in regular self-care activities that promote mental well-being, such as exercise, relaxation techniques, or hobbies you enjoy. Stay connected with your support network, whether it's friends, family, or online communities, and don't hesitate to reach out for help if you need it. Remember that therapy is a valuable tool, and you can always return for

Figure 4.9 Recommendation Dataset

Our Recommendation dataset consists of all the set of recommendations for various combinations possible from the user entry from the UI and suggesting him recommendations based on their particular selection. T-5 mini model summarized the overall recommendation for various selections made by the user.

V. CONCLUSION

We are continuously putting efforts towards improving the recommendation system's accuracy, usability, and performance [1]. This includes those of refining the recommendation algorithm, expanding the dataset for training purposes, integrating advanced NLP techniques, and optimizing the GUI's features based on user feedback adding on to this a robust error-handling mechanism can also be implemented to ensure smooth system operation. From this model, the performance metrics feats achieved refer figure 4.8 can give the idea of its accuracy furthermore and further improvements on this can be added upon. This research aims to contribute to the advancement of recommendation systems, with a focus on delivering an intuitive and effective solution that meets user needs and expectations. Last but not least this model can be built into an app on which after a patient's survey they can get in contact with a certified psychiatrist to further assistance in extreme cases too by maintaining a proper backend and API along with security for additional peace of the user's mind.

VI. REFERENCES

[1] M. Habib, Z. Wang, S. Qiu, H. Zhao and A. S. Murthy, "Machine Learning Based Healthcare System for Investigating the Association Between Depression and Quality of Life," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 5, pp. 2008-2019, May 2022, doi: 10.1109/JBHI.2022.3140433.

[2] S. Ledesma, M. A. Ibarra-Manzano, E. Cabal-Yepez, D. L. Almanza-Ojeda, and J.G.Avina-Cervantes, "Analysis of data sets with learning conflicts for machine learning," IEEE Access, vol.6, no.1, pp.45062–45070,018,doi: 10.1109/ACCESS.2018.2865135.

[3] Centers for Disease Control and Prevention, "National health and nutrition examination survey: Analytic guidelines, 2011–2017," pp. 1.40,2018,[Online].Available:https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx

[4] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2nd ed.,New York, NY, USA: Wiley, 2001.

[5] Passos, Ives Cavalcante, Pedro Ballester, Jairo Vinicius Pinto, Benson Mwangi, and Flávio Kapczinski. "Big data and machine learning meet the health sciences." Personalized psychiatry: big data analytics in mental health (2019): 1-13.

[6] Narayanrao, Purude Vaishali, and P. Lalitha Surya Kumari. "Analysis of machine learning algorithms for predicting depression." In 2020 international conference on computer science, engineering and applications (iccsea), pp.14.IEEE,2020.doi:10.1109/ICCSEA491-43.2020.9132963

[7] V. N. Vapnik, Statistical Learning Theory. New York, NY, USA: Wiley,1998.

[8] C. M. McKercher, M. D. Schmidt, K. A. Sanderson, G. C. Patton, T.Dwyer, and A. J. Venn, "Physical activity and depression in young adults," Amer. J. Preventive Med., vol. 36, no. 2, pp.161–164,2009,doi: 10.1016/j.amepre.2008.09.036.

[9] Gao D, Zhang YX, Zhao YH. Random forest algorithm for classification of multiwavelength data. Research in Astronomy and Astrophysics. 2009 Feb 1; 9(2):220. Doi: 10.1088/1674-4527/9/2/011

[10] M. M. Alsubaie, H. J. Stain, L. A. D. Webster & R. Wadman (2019) The role of sources of social support on depression and quality of life for university student International Journal of Adolescence and Youth, 24:4, 484-496, DOI: 10.1080/02673843.2019.156888

[11] Kolenik, Tine, and Matjaž Gams. "Persuasive technology for mental health: One step closer to (Mental health care) equality?" IEEE Technology and Society Magazine 40, no. 1(2021):80-86. doi: 10.1109/MTS.2021.3056288.

[12] L. Castelli, R. Torta, A. Mussa, P. Caldera, and L. Binaschi, “Fast Screening of depression in cancer patients: The effectiveness of the HADS,” Eur. J. Cancer Care, vol. 20, no. 4, pp. 528–533,2010,doi:10.1111/j.1365-2354. 2010.01217.x

[13] Shatte, A. B. "hutchinson dM, Teague SJ." Machine learning in mental health: a scoping review of methods and applications. Psychol Med 49 (2019):1426-48. doi:10.1017 /S003329171 9000151

[14] Aleem, Shumaila, Noor ul Huda, Rashid Amin, Samina Khalid, Sultan S. Alshamrani, and Abdullah Alshehri. "Machine learning algorithms for depression: diagnosis, insights, and research directions." Electronics 11, no. 7 (2022):1111.doi:org/10.3390/electronics110711

[15] K. P. Linthicum, K. M. Schafer, and J. D. Ribeiro, “Machine learning in suicide science: Applications and ethics,” Behavioral Sci. Law, vol. 37,no. 3, pp. 214–222, Sep. 2018, doi: 10.1002/bsl.2392

[16] M. L. Giger, “Machine learning in medical imaging,” J. Amer. College Radiol., vol. 15, no. 3,pp.512–520,2018,doi: 10.1016/ j.jacr. 2017. 12.028

[17] Ahmad, Roslee, Azizan Mohd Isa, Mohd Radzniwan A. Rashid, Nasir Yusoff, Lokman Mohd Tahir, Miftachul Huda, and Rahmiwati Marsinun. "Prevalence of Anxiety and Depression among Caregivers of Cancer Patients A Case Study in a Public Hospital in Malaysia." IIUM Medical Journal Malaysia 22, no. 1 (2023). Doi:10.31436/imjm.v22i1.2146

[18] Cho, Yunji, Joo Kyung Lee, Do-Hoon Kim, Joo-Hyun Park, Moonyoung Choi, Hyun-Jin Kim, Myung-Ji Nam, Kang-Uk Lee, Kyungdo Han, and Yong-Gyu Park. "Factors associated with quality of life in patients with depression: A nationwide population based study." PloS one 14, no. 7 (2019):e0219455. doi:10.1371/ journal.pone.0219455

[19] Mathers, C. D., and D. Loncar, “Projections of global mortality and burden of disease from 2002 to 2030,” PLoS Med., vol. 3, no. 11, pp. 2011–2030, 2006, doi: 10.1371

[20] S. Senn, M. Tlachac, R. Flores and E. Rundensteiner, "Ensembles of BERT for Depression Classification," 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 4691-4694, doi: 10.1109/EMBC 48229.2022.9871120.