International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Tools and Machine Learning Algorithms for Predicting Depression, Anxiety, and Stress: A Literature Review

[1]Fathima Kondeth, [2]Dr. S Prabhu

[1]*Research Scholar, PG and Research Department, Department of Computer Science, Park's College,* Tirupur, India

[2]*Research Supervisor, PG and Research Department, Department of Computer Science, Government Arts & Science College, Thittamalai,* Tirupur, India

ABSTRACT—

Technology development in this period has made it necessary for the medical industry to evolve. So, it is essential to anticipate and examine the various psychological problems in order to discover a cure as soon as feasible. Machine learning techniques have improved, making it possible to anticipate psychological disorders. This study provides an overview of several psychological diseases brought on by ADHD, bipolar disorder, and mental difficulties following Covid-19. The review considered machine learning techniques for predicting mental health issues. In this review, publications from online databases between 2020 and 2023 were taken into account. A systematic review is conducted using the PRISMA approach. We gathered 96 papers for this study, of which 21 were included following screening and identification. Numerous machine learning techniques have been discovered to be used for prediction. ANN, and CNN. SVM, Logistic Regression, and Random Forest were the most frequently employed algorithms by academics. The datasets used include the National Health and Nutrition Examination Survey, the Open-Source Internet Survey dataset, the Online Questionnaire, the Norway Dataset, the Emotional Recall Data, the Free Associations, the Suicide Notes, and the Valence-Arousal Norms. Also, GAD-7, DASS-21, and DASS-42 were used as tools. The majority of them were discovered to have used DASS-21.

Keywords— Depression, anxiety, stress, prediction, machine learning, DASS-21

Introduction

World Health Organization defines mental health as "a state of mental well-being that enables people to cope with the stresses of life, to realize their abilities, to learn well and work well and to contribute to their communities" [1]. Mood swings, stress, anxiety, and sadness are examples of mental health disorders. There are numerous biological, social, and psychological causes of mental health. Everybody's life is significantly impacted by their mental health. Mental health is as vital, even if it is intangible and, in contrast to physical health, is frequently taken far more seriously. Nonetheless, most individuals opt to disregard it. One cause of this misinformation is the stigmatisation of mental health, which prevents individuals from discussing it and, as a result, lowers their mental state. Due to the stigma associated with mental illness, people who suffer from it often keep it a secret from their families, co-workers, and the general public. Distress and emotional restlessness are brought on by the dysfunctional thought and behaviour patterns that result from mental health illnesses.

A mood disorder having two extremes - depressed ("low") and manic ("high") - is known as bipolar disorder, formerly known as manic depressive illness. The severity varies, and mild cases may seem typical for years. There are many different symptoms; a person may be either manic or predominantly depressive. A person is most likely healthy and able to operate between bouts. Among other signs of depression, a person may feel suicidal, hopeless, and perpetually down. While manic, a person exhibits excessive elation, is more irritable, needs less sleep, develops lofty plans, and may act rashly in ways that could be dangerous. Those who are more stable can benefit from psychotherapy by receiving assistance in identifying and managing their symptoms. Treatment for acute episodes and relapse prevention are both possible with medication. A crucial part of treatment is psychosocial assistance. A mental disease called bipolar disorder includes numerous stages, ranging from mild melancholy to hyperactive episodes. Clinical and research data demonstrate that postponing therapy increases stress levels, causes mood swings, and worsens conditions like cardiac arrest, stroke, heart attack, and depression [2]. The main effects of delayed diagnosis or treatment are high rates of suicide, decreased productivity, and decreased quality of life. It is feasible to forecast bipolar disorder using EEG data.

Luján et al. [3] developed a lasso model that, based on early improvements in ADHD symptoms, may forecast an adult population's response to viloxazine extended release. A thorough examination into the risk factors for children with ADHD was reported by Faraone et al. [4]. The two main goals of the study were to (i) forecast children with ADHD and (ii) look at the risk factors for children with ADHD. Due to behavioural symptom similarities, the present diagnostic procedure for ADHD is time-consuming and challenging. Random Forest was chosen out of about eight algorithms because it was effective at predicting ADHD. There has been a rise in interest in automatic diagnosis of ADHD using machine learning analysis of brain signals. The objective of this study was to develop a reliable model for pattern-based discrimination between ADHD patients and healthy controls [5]. The outcomes showed that the performance of the prediction models can be greatly enhanced by the inclusion of complementing features.

Globally, changed lifestyles have resulted in a major rise in anxiety levels, which in turn leads to stress and depression [6]. Many studies have been conducted in this area. The development of machine learning algorithms makes it simple to predict mental disease in its early stages. Several methods and instruments have been employed for this. A thorough examination of the literature was undertaken to explore the different machine learning tools and methods.

The paper is given in the following order. First, Section 2 gives a brief overview of the methodology used for literature review. Section 3 includes the literature analysis and findings. Discussions on the findings are given in Section 4. The paper ends with Section 5 having conclusion and directions of future work.

METHODOLOGY

In order to present an overview of the developments in research on the diagnosis of mental illnesses for this review, a number of ideas and questions were taken into account during the selection, extraction, and analysis of previous studies. The following issues have been taken into consideration: "what are the tools used for prediction?", "prediction of illnesses that caused mental stress and depression", "Covid-19 triggered depression prediction", "how machine learning can be used for the diagnosis of mental illness", and "How can mental illness be predicted?". Google Scholar, Elsevier ScienceDirect, and IEEE Xplore were the databases that were searched. The terms "depression prediction", "mental health prediction", "machine learning for mental health diagnosis", "artificial intelligence for mental health illness", and "anxiety, depression, and stress prediction" were used to thoroughly search these three well-known electronic databases. The keywords were grouped together using the Boolean AND and OR expressions.

Selection of Papers

The papers were examined in light of their importance to the aforementioned topics. To determine whether they fall under our area of interest, we examined the titles, abstracts, and keywords. Following that, the papers were divided into two groups based on the inclusion and exclusion criteria listed below.

E1: Articles that do not explicitly address the prediction of stress, anxiety, and depression using machine learning methods.

E2: Papers that have little bearing on the issues.

E3: Articles that don't include at least one of the search terms.

E4: Duplicate papers

I1: Full-text articles

I2: English papers

I3: Articles released in 2018 to 2023

I4: Articles on ADHD, the psychological effects of COVID-19, bipolar disorder and the prediction of them.

PRISMA

The Preferred reporting items for systematic review and meta-analysis protocols (PRISMA) protocol [7], which outlines the steps for conducting a systematic review, is the foundation of the study. A total of 96 documents were initially gathered from all internet resources. Due to overlap caused by the Google Scholar search, duplicate papers were eliminated, leaving 72 papers. After reading each paper's title and abstract, 51 irrelevant papers were then eliminated. The final 21 papers were thoroughly examined and included for this review. The screening procedure is shown in Figure 1.

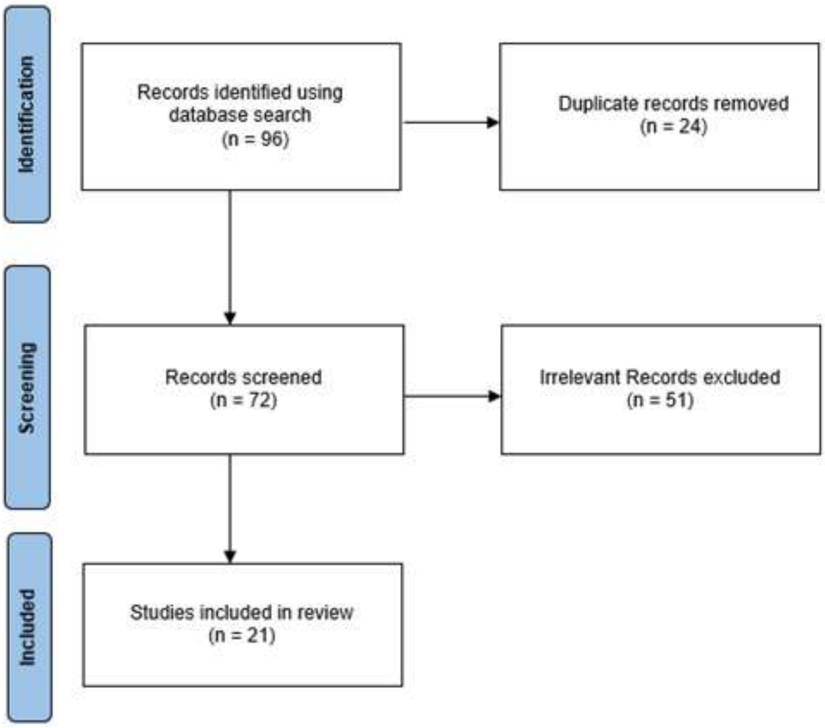

Fig. 1 PRISMA Model

LITERATURE ANALYSIS AND FINDINGS

Fatima et al. [8] used a semi-supervised machine learning model that was trained and tested using four datasets: the Emotional Recall Data (ERT), the Little World of Words English free association dataset, the corpus of real suicide notes, and valence-arousal norms. The ERT dataset consists of 200 people's emotional recalls that have been compared to psychometric scales like the Depression Anxiety Stress Scales (DASS-21). Through the use of a machine learning technique, it is now able to map individual recall of emotive word sequences against their mental wellbeing using a completely anonymous dataset. The Little World of Words project is a global effort to map human semantic memory through free associations, or conceptual links in which the remembrance of one word prompts the recall of others. A wide range of cognitive processes have been successfully predicted using cognitive networks comprised of free associations between concepts. The valence-arousal norms employed here show how exciting/inhibiting (arousal) and pleasant/unpleasant (valence) words are when identified singly within a psychology mega-study. Almost 20,000 English words' valence and arousal norms were included in this dataset. They tested algorithms (i) decision tree, (ii) multilayer perceptron (MLP), and (iii) recurrent neural network (Long–Short Term Memory [LSTM]). It was found that semantic distance for recalls was key for estimating depression levels, but redundant for anxiety and stress levels.

The study by Uddin et al. [9] suggests a practical method for locating texts representing one's own self-reported depressive symptoms using recurrent neural networks (RNNs) based on long short-term memory (LSTM). Then, robust elements that were derived from the reflection of potential depressive symptoms that were previously established by medical and psychological professionals are shown. The characteristics outperform traditional strategies, which are mostly based on word frequencies rather than symptoms. The time-sequential characteristics are then trained to distinguish writings reporting depression symptoms from postings without such descriptions using a deep learning approach (RNN) (non-depression posts). Lastly, depressive posts are automatically predicted using the trained RNN. The system's performance is compared to that of conventional techniques, where it outperformed them all.

Using datasets from the National Health and Nutrition Examination Survey (NHANES), a model is trained by Lee and Kim [10]. The NHANES is a recurrent cross-sectional survey used to track trends in the health and nutritional status of the US population who live in the community and are not institutionalised. With 10-fold cross-validation, six distinct ML classification techniques—artificial neural network, random forest, AdaBoost, stochastic gradient boosting, XGBoost, and support vector machine—were used to classify the data. The best AUC and specificity for predicting depression were attained by an artificial neural network trained with specific features across all classification models. The highest F1-score, sensitivity, accuracy, and precision were achieved by the support vector machine when predicting depression.

In an effort to address the majority of COVID-19-related worries, Joshi et al. [11] conducted a survey. These worries included adapting to a stay-at-home routine, its effects on relationships, irritability brought on by working from home, difficulty juggling work and household responsibilities, familiarity with online classes, difficulty focusing online, ability to connect with others, desire to work during the pandemic, inability to participate in pre-pandemic activities, and financial instability. Naive Bayes, Random Forest, Logistic Regression, Support Vector Machine, and K-nearest Neighbor are the five methods that are applied. SVM was discovered to have the best performance out of the five.

Bagga et al. [12] discovered online communities of depressed persons who post their ideas, feelings, and support in these forums. By manipulating the data, extracting features, categorising, and attempting to understand what the characteristics of "depressed" text are, we will analyse the "depressed" text in this work and attempt to "predict" whether a text should be classified as depressed or not. Three distinct machine learning algorithms were employed to predict depression using text analysis and text data mining techniques on the text extracted from social forums: Multinomial Naive Bayes, SVC, and K-nearest Neighbors. Positive sentiment, neutral sentiment, and negative sentiment were the three emotional state indicators taken into account. Text mining methods have been used for information retrieval. Positive and negative sentiments have been classified as the polar opposites of each post's sentiment.

A cross-sectional study was carried out in Saudi Arabia during the COVID-19 pandemic to gauge the prevalence of generalised anxiety disorder (GAD) [13]. The GAD-7 was adopted by the researchers. The J48 Decision Tree was used to create the prediction models due to its interpretability and comprehensibility and the Support Vector Machine classifier's robust results in data pertaining to medicine. The early classification of two-class and three-class anxiety disorders produced positive experimental findings. In terms of performance, the Support Vector Machine classifier performed better than the J48 Decision Tree.

Ryu et al. [14] conducted a study on 31 patients who had their first stroke and were identified as having PSD at the time of admission. All PSD patients received psychological and medical care for 4 weeks after their diagnosis, including antidepressants (i.e., escitalopram, amitriptyline, or fluoxetine). The cognitive and functional changes in controls—a total of 35 age-matched patients who experienced their first stroke without PSD—were compared to those in PSD patients. Support vector machines, k-nearest neighbours, random forests, voting ensemble models, and statistical analysis utilising logistic regression were used as machine learning techniques. With the aid of a support vector machine linear algorithm, PSD was effectively predicted.

Atlam et al. [15] used a dataset to investigate the psychosomatic effects of the digital learning tools used in COVID-19's online courses on students' academic performance in universities in the Arab world. The information was acquired with a specific focus on the psychosomatic effects of online learning tools before and after COVID-19 and consists of five key components: (1) Digital gadgets including laptops, smartphones, and iPads; (2) Sleeping patterns; (3) Social Engagement; (4) Mental State; and (5) Academic Performance. These machine learning models' performances—Logistic Regression, Random Forest, Decision Tree, XGB, AdaBoost, Deep Learning, and SVC—have all been taken into account. Regression using logs fared better.

Qasrawi et al. [16] gathered data through a multidisciplinary study on the factors affecting students' physical, mental, and social health in the West Bank and East Jerusalem. The study used a 3984-student representative sample drawn from 100 schools. The study evaluated the performance and accuracy of 5 ML models in predicting the associated health indicators on the sadness and anxiety of Palestinian schoolchildren. The study's findings support the notion that ML algorithms, particularly RF and neural networks, are useful predictors of students' mental health. Public health experts, medical practitioners, and decision-makers will be able to foresee emerging problems and develop pertinent intervention programmes to improve kids' health, education, and wellbeing by using precise ML techniques like RF.

Priya et al. [17] used data from employed and unemployed individuals with a wide range of responsibilities from household chores to professional activities through a questionnaire. Subsequently the data is categorised using five machine learning techniques, namely Decision Tree, Random Forest Tree, Naïve Bayes, Support Vector Machine and KNN. The accuracy of Naïve Bayes was shown to be the highest, whereas Random Forest was recognised as the best model. The relevant variables were found to be 'scared without any good reason', 'Life was meaningless' and 'Difficult to relax' for the scales of Anxiety, Depression and Stress, respectively. As such, these variables were judged to be most important in recognising psychological disorder.

The DASS-21 online version was used to gather the data by Sanjay et al. [18]. The information was gathered using the online DASS-21 version. KNN, Random Forest, and Logistic Regression were the three machine learning algorithms that were employed. They found that Random Forest performed better than both Logistic Regression and KNN after running the models.

Kumar et al. [19] used eight machine learning algorithms to data gathered from the online DASS42 tool for the prediction of the occurrence of psychological issues like anxiety, depression, and stress. These included Bayes Net, Naive Bayes, Multilayer Perceptron, RBFN, K-Star, K-nearest Neighbor, J48, and a hybrid classification algorithm (K-Star with Random). The same techniques were also used on DASS21, a different dataset that the authors collected. The outcomes demonstrated that neural networks outperformed all others. In both datasets, the neural network category's performance for depression was the best for the RBFN. Yet, the DASS21 anxiety result from random forest is 100%. This happened as a result of the dataset being skewed and only using a small number of instances.

Singh et al. [20] also used the surveys stated by DASS-21. The Random Forest algorithm's accuracy was shown to be superior to that of the decision tree. To determine the most effective model for predicting mental illness, the F1 score was used. The F1 score indicates that Random Forest is the best approach.

Results and Discussion

The datasets, machine learning techniques, and mental diseases used by various researchers are listed in Table 1. While some researchers have examined depression, anxiety, and stress all at once, others have focused solely on one of these. They have employed a variety of techniques, including online questionnaires, to gather data. The main tool was DASS-21. Also, the samples varied, including those who were in school, individuals affected by Covid-19, employed, and unemployed people. The table also includes a list of the researchers' algorithms.

DATASETS AND ML ALGORITHMS

Paper	***Year***	***Mental Illness***	***Dataset Used***	***ML Algorithm Used***
[8]	2021	Depression, anxiety, stress	ERT, the Little World of Words English free association dataset, the corpus of real suicide notes, and valence-arousal norms	Decision tree, multilayer perceptron (MLP), recurrent neural network (LSTM)
[9]	2021	Depression	Text based data	RNN, LSTM
[10]	2022	Depression	Health and nutrition examination survey	Artificial neural network, random forest, AdaBoost, stochastic gradient boosting, XGBoost, support vector machine
[11]	2022	Depression	Survey	Naive Bayes, Random Forest, Logistic Regression, Support Vector Machine, and K-nearest Neighbor
[12]	2021	Depression	Text mining using NLP	Multinomial Naive Bayes, SVC, and K-nearest Neighbors
[13]	2022	Anxiety	Online survey (GAD-7)	Support Vector Machine, J48 Decision Tree
[14]	2022	Depression	Data from poststroke depression patients	Support vector machines, k-nearest neighbors, random forests, voting ensemble models, logistic regression
[15]	2022	Psychological impact and learning difficulties	Online questionnaire	Logistic Regression, Random Forest, Decision Tree, XGB, AdaBoost, Deep Learning, SVC
[16]	2022	Anxiety, depression	Data from school students	Random forest, neural network, decision tree, SVM, Naïve Bayes
[17]	2020	Depression, anxiety, stress	Data from employed and unemployed individuals (DASS-21)	Decision Tree, Random Forest Tree, Naïve Bayes, Support Vector Machine, KNN
[18]	2022	Depression, anxiety, stress	DASS-21 online	KNN, Random Forest, and Logistic Regression
[19]	2020	Depression, anxiety, stress	Online DASS42, DASS-21	Bayes Net, Naive Bayes, Multilayer Perceptron, RBFN, K-Star, K-nearest Neighbor, J48, hybrid classification algorithm (K-Star with Random)
[20]	2022	Depression, anxiety, stress	DASS-21	Random Forest, decision tree

A number of machine learning algorithms were found in the analysed papers that were used by different researchers. These included Decision Tree, J48 Decision Tree, K-nearest Neighbor, K-Star, Logistic Regression, Multilayer Perceptron, Naive Bayes, Neural network, Random Forest, Radial Basis Function Network, Recurrent neural network (Long Short Term Memory networks), Stochastic gradient boosting, Support Vector Machine, Support Vector Classification, Voting ensemble models and XGBoost. Figure 2 shows the occurrences of mostly used machine learning algorithms found in the papers examined.

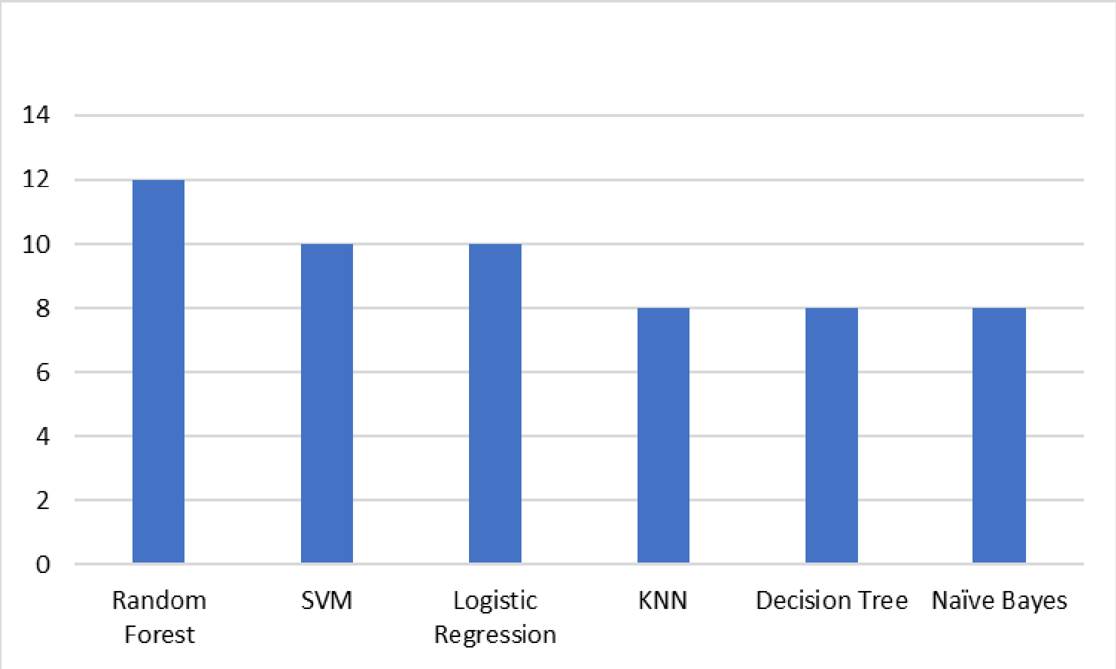

Fig.2 Mostly used ML Algorithms

Conclusion and Future Work

This research examines the use of machine learning approaches for mental disorder analysis and prediction. Also, the application and development of machine learning techniques over time have been documented. When used on different datasets, different machine learning approaches perform in different ways. It has been noted that machine learning approaches utilised in these study publications have demonstrated remarkable results in the prediction and analysis of psychological diseases all over the world, regardless of frequency and differences in performance. The majority of machine learning algorithms might not perform equally well on all problems. Depending on the data samples obtained and the data's properties, the machine learning models' performance will change. To provide good data that could enhance clinical practise and decision-making, challenges and limitations experienced by the researchers must be well controlled. As they produced better results, the majority of them utilised techniques like SVM, Random Forest, and Logistic Regression. DASS-21 was the main tool employed by the researchers. For more accurate mental illness prediction, hybrid models with several machine learning algorithms incorporating data from DASS-21 surveys can be investigated.

References

World Health Organization, “World Mental Health Report: Transforming Mental Health for all.” 2022.

N. Agnihotri and S. K. Prasad, “Bipolar Disorder: Early Prediction and Risk Analysis using Machine Learning,” vol. 21, no. 8, 2022.

M. Á. Luján, A. M. Torres, A. L. Borja, J. L. Santos, and J. M. Sotos, “High-Precise Bipolar Disorder Detection by Using Radial Basis Functions Based Neural Network,” Electronics, vol. 11, no. 3, p. 343, Jan. 2022, doi: 10.3390/electronics11030343.

S. V. Faraone et al., “Predicting efficacy of viloxazine extended-release treatment in adults with ADHD using an early change in ADHD symptoms: Machine learning Post Hoc analysis of a phase 3 clinical trial,” Psychiatry Research, vol. 318, p. 114922, Dec. 2022, doi: 10.1016/j.psychres.2022.114922.

Md. Maniruzzaman, J. Shin, and Md. A. M. Hasan, “Predicting Children with ADHD Using Behavioral Activity: A Machine Learning Analysis,” Applied Sciences, vol. 12, no. 5, p. 2737, Mar. 2022, doi: 10.3390/app12052737.

E. Ghasemi, M. Ebrahimi, and E. Ebrahimie, “Machine learning models effectively distinguish attention-deficit/hyperactivity disorder using event-related potentials,” Cogn Neurodyn, vol. 16, no. 6, pp. 1335–1349, Dec. 2022, doi: 10.1007/s11571-021-09746-2.

D. Moher et al., “Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement,” PLoS Med, vol. 6, no. 7, Jul 2009, doi: 10.1371/journal.pmed.1000097

A. Fatima, Y. Li, T. T. Hills, and M. Stella, “DASentimental: Detecting Depression, Anxiety, and Stress in Texts via Emotional Recall, Cognitive Networks, and Machine Learning,” BDCC, vol. 5, no. 4, p. 77, Dec. 2021, doi: 10.3390/bdcc5040077.

M. Z. Uddin, K. K. Dysthe, A. Følstad, and P. B. Brandtzaeg, “Deep learning for prediction of depressive symptoms in a large textual dataset,” Neural Comput & Applic, vol. 34, no. 1, pp. 721–744, Jan. 2022, doi: 10.1007/s00521-021-06426-4.

C. Lee and H. Kim, "Machine learning-based predictive modeling of depression in hypertensive populations," PLoS ONE, vol. 17, no. 7, p. e0272330, Jul. 2022, doi: 10.1371/journal.pone.0272330.

S. Joshi, A. Ghosh, and S. Shukla, "Machine Learning Model for Depression Prediction during COVID-19 Pandemic," in 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, Apr. 2022, pp. 1–5. doi: 10.1109/ICDCECE53908.2022.9792792.

N. Bagga, P. Vashistha, and P. Yadav, "Predicting Depression from Social Networking Data using Machine Learning Techniques," in 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, Dec. 2021, pp. 128–132. doi: 10.1109/ICAC3N53548.2021.9725402.

F. M. Albagmi, A. Alansari, D. S. Al Shawan, H. Y. AlNujaidi, and S. O. Olatunji, "Prediction of generalized anxiety levels during the Covid-19 pandemic: A machine learning-based modeling approach," Informatics in Medicine Unlocked, vol. 28, p. 100854, 2022, doi: 10.1016/j.imu.2022.100854.

Y. H. Ryu et al., "Prediction of Poststroke Depression Based on the Outcomes of Machine Learning Algorithms," JCM, vol. 11, no. 8, p. 2264, Apr. 2022, doi: 10.3390/jcm11082264.

E.-S. Atlam, A. Ewis, M. M. A. El-Raouf, O. Ghoneim, and I. Gad, "A new approach in identifying the psychological impact of COVID-19 on university student's academic performance," Alexandria Engineering Journal, vol. 61, no. 7, pp. 5223–5233, Jul. 2022, doi: 10.1016/j.aej.2021.10.046.

R. Qasrawi, S. P. Vicuna Polo, D. Abu Al-Halawa, S. Hallaq, and Z. Abdeen, "Assessment and Prediction of Depression and Anxiety Risk Factors in Schoolchildren: Machine Learning Techniques Performance Analysis," JMIR Form Res, vol. 6, no. 8, p. e32736, Aug. 2022, doi: 10.2196/32736.

A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," Procedia Computer Science, vol. 167, pp. 1258–1267, 2020, doi: 10.1016/j.procs.2020.03.442.

V. M. Sanjay, A. I, D. Gc, A. Hp, J. Malik, and T. Bn, "Anxiety Prediction during Stressful Scenarios using Machine Learning," in 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, May 2022, pp. 1199–1202. doi: 10.1109/ICICCS53718.2022.9788151.

P. Kumar, S. Garg, and A. Garg, "Assessment of Anxiety, Depression and Stress using Machine Learning Models," Procedia Computer Science, vol. 171, pp. 1989–1998, 2020, doi: 10.1016/j.procs.2020.04.213.

P. Singh, G. Singh, and S. Bharti, "The Predictive Model of Mental Illness using Decision Tree and Random Forest classification in Machine Learning," in 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, Apr. 2022, pp. 01–05. doi: 10.1109/ICACITE53722.2022.9823761.