

Early Detection of Anxiety, Depression and Stress among Potential Patients using machine learning and deep learning models

Alphonsa Sini P.J
Research Scholar
APJ Abdul Kalam Technological University
Kerala, India
alphonsasini@gmail.com

Dr Sherly K.K
Associate Professor
Rajagiri School of Engineering And Technology
Kerala, India
sherlykk@rajagiritech.edu.in

Abstract— Depression, anxiety and stress are the three important mental health conditions. Only a small portion of the millions of people who experience depression, anxiety and stress each year receives a prompt treatment. The development of technology that helps clinicians detect various types of mental diseases has greatly benefited from the progress of Machine Learning (ML) techniques over the past few decades. This study presents a comparison of three important mental health issues (Depression, Anxiety, and stress) using machine learning and Deep Learning (DL) algorithms. The proposed work also investigates about how the ML and DL techniques might help with the detection, diagnosis, and management of mental health diseases with a better effective and novel approach. As a result, three different algorithms (SVM, ANN and XgBoost) has been implemented in which Support vector machine (SVM) has achieved the better accuracy in detecting the mental health. The accuracy achieved in detecting depression, anxiety and stress are 99.32%, 99.80% and 98.44% respectively. Henceforth the proposed study was able to achieve a satisfactory result that helps to improve the decision-making and clinical practice.

Keywords—Mental health issues, Depression, Anxiety, Stress, psychiatric issues

I. INTRODUCTION

People's minds are psychologically impacted by the contemporary day lifestyle, which leads to depression and mental anguish. Depression is a persistent mental disorder that affects how someone thinks and develops mentally. Suicidal thoughts are more common in those with depression [1]. Every year, almost 800,000 people die by suicide. To address the stress of mental health disorders, a holistic approach is necessary. The socioeconomic standing of an individual may suffer from depression. People with depression are less inclined to socialize. Depression can be fought through counseling and psychiatric treatments [2].

Algorithms with the capacity to teach the users to recognize intricate patterns are the goal of machine learning (ML). Through the use of prior information and answers, this capacity aids in solving new challenges. The implementation of ML algorithms produces outcomes that are regulated and standardized [3]. The major inputs used by supervised machine learning algorithms are used to identify known values, whereas the key inputs used by unsupervised machine learning algorithms reveal unrecognized structures and clusters within the data collected [4].

The applicability of applying ML approaches in the healthcare industry has been demonstrated by their capacity to analyses massive amounts of diverse data and delivers insightful clinical information. ML-based methods help mental health professionals make predictions and understand mental illnesses more effectively [5].

By extracting knowledge from unstructured medical data, machine learning techniques aid in healthcare prediction and diagnosis. The prediction results assist in identifying high-risk medical issues in patients so that they can receive treatment sooner. Healthcare professionals can predict the occurrence of mental disorders and provide appropriate treatment outcomes with the use of machine learning (ML) algorithms that help evaluate probable behavioral biomarkers in mental disorders. The methods aid in the visualization and comprehension of challenging medical data [6].

The visualization aids in creating a strong theory for mental condition diagnosis. The intricacy of depression is not correctly identified by the conventional clinical diagnostic method. By employing ML approaches, it is simple to identify and foresee the makeup of the symptoms associated with mental diseases like depression [7]. As a result, it appears that the ML-based diagnostic technique is an effective one for predictive analysis. In the healthcare industry, sensor, text, structured data, and multimodal technological interactions are the key domains used for extracting observations linked to mental diseases through machine learning (ML). With the aid of mobile devices and audio signals, the sensor data can be evaluated [8].

Clinical records, text messages, and social media sites can all be used to extract text sources. The structured data are data that have been taken out of common screening tools, surveys, and health records. The data from human interactions with common technological devices, robots, and virtual agents are included in the multimodal technology interactions. The ML techniques can help with the diagnosis of mental health issues. For diagnosing mood problems, the bulk of studies examine Twitter data and sensor data from mobile devices [9].

Nearly one-third of the population suffers from anxiety disorders at some point during their lives. Anxiety disorders have a major negative impact on psychological functioning, workplace productivity, and overall quality of

life, as well as a high financial cost. Despite the fact that there are effective treatments for anxiety disorders, more than one-third of persons who are diagnosed with the problem are not offered treatment [10].

Early diagnosis of mental health issues is critical for a better understanding of mental diseases and better medical care. Artificial intelligence (AI) and machine learning (ML) approaches are being used to increase the understanding of mental health problems and to aid mental health therapists in making better therapeutic judgments because to the large amount of evidence on an individual's mental health status. Some indicators may be identified if a pattern among people with these diseases is investigated. While these patterns aren't visible to the naked eye, they can be detected using machine learning approaches based on previously gathered data.

Anxiety disorders can cause depression since they are so stressful and interfere with a person's life. Mental disorders can develop in people which may lead to risk their lives. It is important to predict the mental illnesses in patients at the proper time in order to avoid the consequences. Machine learning algorithms are mostly used to detect the disorder of the anxiety illness, as well as predict risk levels and treatment response levels. Hence the aim of this study is to detect Anxiety, Depression and Stress among potential patients using machine learning and deep learning models utilizing a suitable dataset containing details of patient suffering from mental health issues.

II. RELATED WORKS

Anxiety disorders and related characteristics have been increasingly relevant in the study of physical health and health behaviors over time. Overuse and underuse of healthcare can have major effects in both the personal and public health domains. Anxiety disorders are linked to reassurance-seeking, avoidance behaviors, and other personality traits that may influence healthcare usage patterns. One health-related activity that has get a lot of attention in relation to anxiety is healthcare utilization, or engagement with a medical care provider or facility [11].

Traditional Machine Learning approaches like linear regression, support vector machines, and random forests are simple to define in terms of which attributes have the biggest impact on the outcome, but they don't account for the time dimension by default. One method for transferring the sequence of medical history to these models is to use temporal patterns. The majority of contributions to the field focus on discovering and finding frequent patterns in massive datasets, then evaluating their performance in tasks such as prediction and categorization [12]. There are a variety of methods for converting one representation to another.

Researchers Fairbrother et al., [13] stated that clinicians are aware of the importance of perinatal AD screening, and there has been a request for recommendations on the best screening methods. The examination of screening methods for prenatal anxiety and associated problems is one more in a limited but expanding

list of studies. The research suggests that methods that have not been created expressly for the assessment of the whole spectrum of AD are unlikely to function at a level good enough for broad use. However, the implementation should have evaluated the essential characteristics of main anxiety disorders and associated disorders in order to have the potential to deliver very accurate screening.

The authors Islam et al., [14] suggested that there are high rates of panic and generalized anxiety in Bangladesh, and that these conditions are linked to particular socio-demographic characteristics, suggesting that during this COVID-19 outbreak, specialized mental healthcare services for people may be especially beneficial to these populations in lowering panic and anxiety. Because the dataset was conducted only in online, it had the drawback of not being representative. Additionally, because it is cross-sectional in nature, this study can only establish estimates and correlations of panic and anxiety, not their possible long-term effects.

The authors Sau and Bhakta [15] proposed a study that focuses on using machine learning technology to automate the detection of mental health issues. With the help of this technology, labor-intensive, manual techniques for assessing anxiety and depression using different rating scales can be substituted with an automated, computer-based method that nevertheless provides reasonable accuracy. Future research could create a computerized automated anxiety and depressive screening system that would be connected to the treating psychiatrist for evaluation and further iterations to improve the system's predicted accuracy.

According to the authors Smys and Jennifer S.Raj [16] the risk of suicide associated with cardiovascular illness can be determined by the severity of the depression and the raised parameter. These signs are broken down into many categories and explained in terms of the psychiatric condition that goes along with anxious depression. Social media is widely used and gives people the ability to express their opinions in order to stay connected for a very long period online [17]. Many characteristics of social media comments and postings were used to train the suggested model. However future research ought to make use of the broader emotional characteristics dataset. To further confirm the efficiency of the suggested system, other datasets should have been employed.

The researchers Bobade and Vani [18] aimed to automatically identify a person's state of stress; the authors used physiological data collected during stressful scenarios. The study has comprehended the structure and format of the publicly accessible WESAD dataset, cleaned and transformed the data to a set eligible to be used in the development of machine learning and deep learning classification methods, explored and created a variety of classification models, and compared them. [19]. The limitation is that the utilized dataset were pre-defined hence the dataset can be gathered via a number of structured questionnaires, which can be done in the future.

The authors Nemani et al.,[20] studied about the relationship between COVID-19-positive people' mortality and psychiatric illnesses. Anxiety disorders, mood disorders, and schizophrenia spectrum disorders were the three categories used to classify psychiatric diseases. The limitation is that if some psychological illnesses are linked to a higher incidence of terminal illness in patients, more investigation is required. To stop health disparities from getting worse, targeted interventions may be required for patients with severe mental illness.

As per the authors Salkevicius et al., [21] physical responses involve the adrenal glands, the sympathetic nervous system, and the release of hormones, while psychological symptoms include irritation, stress, disruption, and restlessness. The symptoms of anxiety disorders, which are a subset of mental illnesses, include concern, social withdrawal, panic attacks, fear, emotional distress, and difficulty falling or staying asleep. Different emotional cues, such as static and dynamic audio/video stimuli, can be used to decipher these symptoms. However, the researchers should have concentrated on incorporating the framework for predicting anxiety levels into the cloud-based system's user interface, making it simple for the psychologist to monitor and respond to variations in the patient's anxiety throughout.

A. Research gap

Depression and other mental health difficulties are frequent, and research has shown that they have a significant impact on physical health. Artificial intelligence (AI) technologies have recently been created to help mental health practitioners such as psychiatrists and counselors make decisions based on evidence presented by patients. Medical history, behavioral data, and social media activity are just a few examples. Additionally, observational research analyzing how communication occurs during bedside contacts, ward rounds, and clinical handovers might be enhanced in order to support patient and family choices through active verbal involvement [18].

Future research should look into how to serve the patients with complex demands, as well as strategies for facilitating engagement through written communication with the patient portal and encrypted messaging [22]. Depression screening equipment is now available, and depressed persons must actively participate. However, a large section of the population refuses to seek professional help owing to a lack of information and the seriousness of the condition. For such a device, a model capable of recognizing depression symptoms from conversational audio is necessary [17].

One of the most common and well-studied psychological effects of mental health is anxiety. An individual's social relationship tends to gather data from their numerous personal digital device and social media interactions, which can be easily mined for behavioral or mental health insights. The failure to detect their mental health in a clear and systematic way was blamed on few methodological difficulties. More recent research has shown substantially less accurate results when using social anxiety

questionnaires and mental diagnostic tests in bigger samples of people who are suffering from anxiety [19]. Henceforth, the proposed study supports to obtain efficient and accurate results in detecting Anxiety, Depression and Stress using machine learning and deep learning models.

III. METHODOLOGY

The proposed study was analyzed using the data which was collected with an on-line version of the Depression Anxiety Stress Scales DASS [23]. For implementing, comparison has been done for the three particular mental health issues (Depression, Anxiety and stress) with three different machine learning and deep learning algorithms (SVM, ANN and XgBoost) along with the use of Random forest which was used for feature selection. Also since the data in the dataset may be imbalanced, SMOTE techniques has been used in order to balance the data and predict the mental health issues easier.

Preprocessing is the process of transforming raw data into something that can be used by a machine learning model. It is the initial and most significant step in developing a deep learning method. The majority of the time, the dataset has duplicate values, missing values, and may even be in an unacceptable format, making it impossible to use machine learning models on it directly. The obtained model accuracy and efficiency are increased through data preparation, which is necessary to clean the data and make it suitable for the model. Hence in the dataset the null values have to be checked and removed.

Exploratory Data Analysis (EDA) is a method for examining datasets to highlight their major characteristics, frequently using visual techniques. Before beginning the modelling work, EDA is used to see what the data speaks about. Finding significant data qualities by looking at a column of numbers or an entire spreadsheet is difficult. Insights can be gained by looking at plain numbers, but doing so can be time-consuming, dull, or overwhelming. Techniques for exploratory data analysis have been developed to help with this challenge.

As the dataset is highly imbalanced the dataset have to be balanced in order to get better models and better predictions to identify attacks. There are two ways to perform operation which results in the balanced data. Over sampling has to be done for the dataset as the lowest target variable is so low and also the other target variable's data should not be lost.

Synthetic Minority Oversampling Technique is the method that will be used to change the dataset (SMOTE). SMOTE picks examples that are close to one another in the feature space, draws a line connecting the examples, and then creates a new sample at a location along the line. In particular, a randomly selected example from the minority class is first picked. Then, for that example, k of the closest neighbors are located (k is equal to 5).

A synthesized example is produced at a randomly chosen location in feature space between the two instances, using a neighbor that was chosen at random as its neighbor.

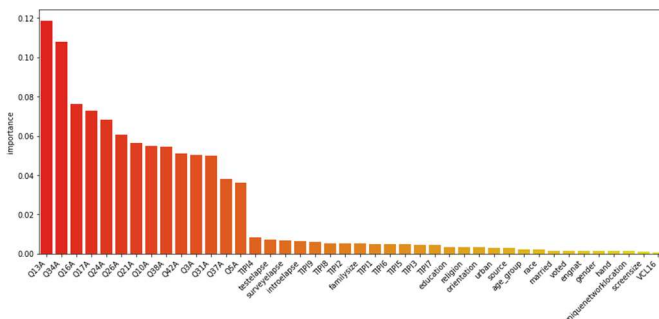
For the minority class, this approach can be used to produce as many synthetic cases as necessary. The paper advocates utilizing SMOTE to oversample the minority class in order to balance the class distribution, followed by employing random under sampling to reduce the number of samples in the majority class. After resampling the dataset using SMOTE technique the minority class is being changed and added more rows to it in order to obtain same distribution of the dataset for all the classes in the target variable.

The data contains the depression target variable in it and will be stored separately since it might be useful for the balance data in model training. Next step is to standardize the data. Standardizing a dataset involves rescaling the distribution of values so that the mean of observed values is 0 and the standard deviation is 1. Standardization assumes that the observations fit a Gaussian distribution (bell curve) with a well-behaved mean and standard deviation and it can be used when the data has input values with differing scales.

Random forest algorithm is used for feature selection. It is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to trying to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. The important features obtained after feature training random forest feature selection are given in the figure 1. Also for feature selection first 15 features which has importance % greater than 0.04 will be considered.

A. Depression detection Process

Performance analysis of depression has been analyzed using SVM, ANN and XGboost with DASS dataset. As there are no highly correlated features there is no need to remove features from the dataset, also all the questions about depression has better correlation than the other towards the depression scores feature can be observed through figure 3 and the target variable is depicted in figure 4. From the analysis, the dataset for all 3 detection (depression, anxiety and stress) contains heavily imbalanced data; henceforth the Smote technique will be utilized in order to balance the data.



B. Anxiety Detection Process

In anxiety detection process, the performance metric has been analyzed using SVM. The accuracy is of 99.98%, while the accuracy for anxiety detection using ANN and XG Boost is 99.2% and 93.3% respectively. It is seen that the SVM has high accuracy.

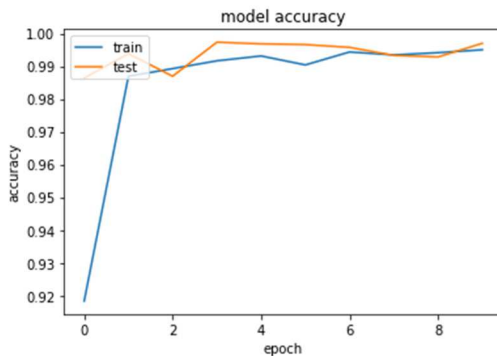


Fig. 3. Accuracy plot

C. Stress Detection process

In Stress detection process, the performance metric has been analyzed using SVM. The accuracy is of 98.44%, while the accuracy for anxiety detection using ANN and XG Boost is 96.82% and 98.40% respectively.

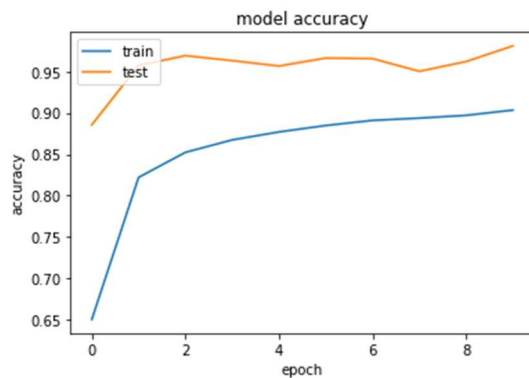


Fig. 4. Model accuracy

D. Testing accuracy comparison

The performance comparison of various measures (accuracy, precision, F1 score, Recall) for all three algorithms (SVM, ANN and Xgboost) in all three mental health issues (Depression, Anxiety and stress) are given in figure 5, figure 6 and figure 7 respectively.

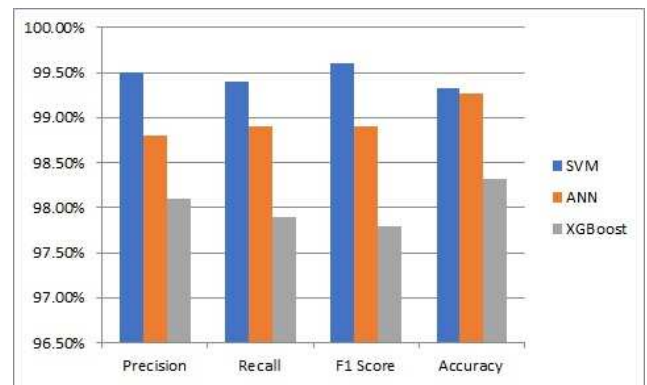


Fig. 5. Performance Comparison of Depression

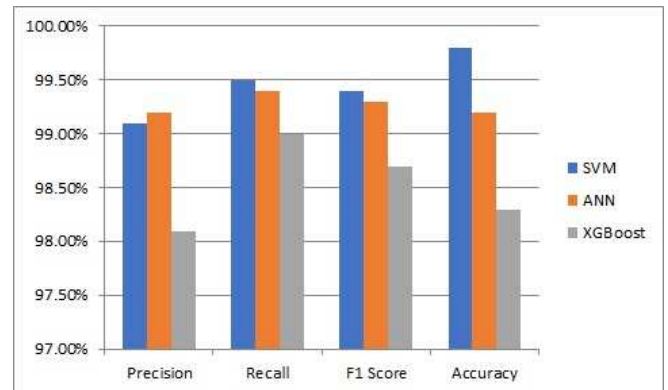


Fig 6. Performance Comparison of Anxiety

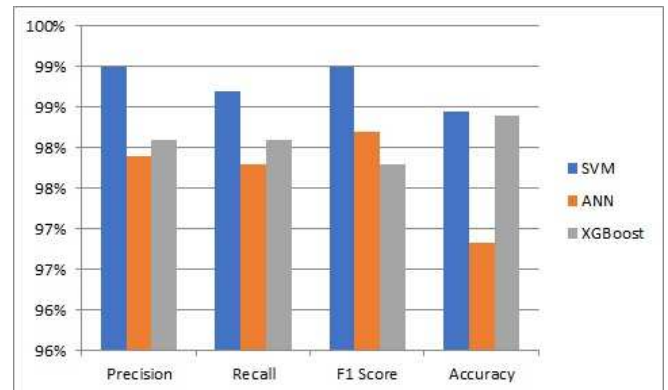


Fig 7. Performance Comparison of Stress

V. CONCLUSION

To assess and address the issues with mental health, several new approaches including algorithms had been established. Numerous solutions can yet be improved. Additionally, there are still a lot of issues on machine learning for such field of mental health that need to be identified and evaluated in a range of situations. The characteristics utilized with in machine learning algorithms should have a substantial impact on the effectiveness of the classification since categorizing the mental healthcare information is often a very difficult job. The studies and research that have already been done indicate whether machine learning could be a helpful tool in understanding psychiatric diseases. In addition, it might aid in identifying

and categorizing patients' mental health issues for future therapy [24].

A practical tool for identifying the emotional state and reactions from patients, among other things, is a more recent technique that makes use of data that results from the integration of multiple sensor modalities found in technologically advanced equipment. Due to a lack of sufficient acceptable verified data, particularly from external sources, it is obvious that the majority of studies and research are still experiencing difficulties validating their findings. In addition, not every machine learning algorithm will perform equally well on every task. The quality of such machine learning models' performance will vary according to the data samples they have access to and the characteristics of the data [25].

Therefore in order to overcome such challenges the proposed study has helped a lot by comparing the three particular mental health issues (Depression, Anxiety and stress) with machine learning and deep learning algorithms. As a result, Support vector machine (SVM) has achieved the better accuracy in which it helps to detect the mental health with more efficient. The accuracy achieved in detecting depression, anxiety and stress are 99.32%, 99.98% and 98.57% respectively. Thus the proposed study was able to achieve satisfactory results that could improve the clinical practice and decision-making.

For future scope, the implementation of the machine learning and deep learning techniques can be extended to other psychiatric issues with different machine learning algorithms. Also it requires comprehensive adoption and computational linguistics in order to increase the accuracy and precision of mental health problem detection in the future. So that It can certainly decrease the amount of individuals who suffer from issues with mental health or whose disorders are thought to go worse without additional therapy.

REFERENCES

- [1] C. Chang *et al.*, "Mental health issues and psychological factors in athletes: Detection, management, effect on performance and prevention: American Medical Society for Sports Medicine Position Statement-Executive Summary," *Br. J. Sports Med.*, vol. 54, no. 4, pp. 216–220, 2020, doi: 10.1136/bjsports-2019-101583.
- [2] S. D'Alfonso, "AI in mental health," *Curr. Opin. Psychol.*, vol. 36, pp. 112–117, Dec. 2020, doi: 10.1016/j.copsyc.2020.04.005.
- [3] R. Yang, J. Curtis, C. Jensen, P. Levy, K. Chown, and J. M. Lappin, "Detection and intervention in emerging youth mental health issues: Outcomes from the first year of the <sc>CASPAR</sc> service," *Early Interv. Psychiatry*, vol. 15, no. 1, pp. 167–173, Feb. 2021, doi: 10.1111/eip.12956.
- [4] D. Ramirez-Cifuentes, C. Largeron, J. Tissier, R. Baeza-Yates, and A. Freire, "Enhanced Word Embedding Variations for the Detection of Substance Abuse and Mental Health Issues on Social Media Writings," *IEEE Access*, vol. 9, pp. 130449–130471, 2021, doi: 10.1109/ACCESS.2021.3112102.
- [5] S. Ji, X. Li, Z. Huang, and E. Cambria, "Suicidal ideation and mental disorder detection with attentive relation networks," *Neural Comput. Appl.*, Jun. 2021, doi: 10.1007/s00521-021-06208-y.
- [6] R. Ahuja and A. Banga, "Mental Stress Detection in University Students using Machine Learning Algorithms," *Procedia Comput. Sci.*, vol. 152, pp. 349–353, 2019, doi: 10.1016/j.procs.2019.05.007.
- [7] D. Ramírez-Cifuentes *et al.*, "Detection of suicidal ideation on social media: Multimodal, relational, and behavioral analysis," *J. Med. Internet Res.*, vol. 22, no. 7, pp. 1–16, 2020, doi: 10.2196/17758.
- [8] D. M. Low, K. H. Bentley, and S. S. Ghosh, "Automated assessment of psychiatric disorders using speech: A systematic review," *Laryngoscope Investig. Otolaryngol.*, vol. 5, no. 1, pp. 96–116, Feb. 2020, doi: 10.1002/lio2.354.
- [9] B. Levis, A. Benedetti, and B. D. Thombs, "Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis," *BMJ*, p. 11476, Apr. 2019, doi: 10.1136/bmj.11476.
- [10] M. G. Wheaton, G. R. Messner, and J. B. Marks, "Intolerance of uncertainty as a factor linking obsessive-compulsive symptoms, health anxiety and concerns about the spread of the novel coronavirus (COVID-19) in the United States," *J. Obsessive. Compuls. Relat. Disord.*, vol. 28, p. 100605, Jan. 2021, doi: 10.1016/j.jocrd.2020.100605.
- [11] M. Taquet, S. Luciano, J. R. Geddes, and P. J. Harrison, "Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA," *The Lancet Psychiatry*, vol. 8, no. 2, pp. 130–140, Feb. 2021, doi: 10.1016/S2215-0366(20)30462-4.
- [12] M. G. Mazza *et al.*, "Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors," *Brain. Behav. Immun.*, vol. 89, pp. 594–600, Oct. 2020, doi: 10.1016/j.bbi.2020.07.037.
- [13] N. Fairbrother, B. Corbyn, D. S. Thordarson, A. Ma, and D. Surm, "Screening for perinatal anxiety disorders: Room to grow," *J. Affect. Disord.*, vol. 250, pp. 363–370, May 2019, doi: 10.1016/j.jad.2019.03.052.

- [14] M. S. Islam, M. Z. Ferdous, and M. N. Potenza, "Panic and generalized anxiety during the COVID-19 pandemic among Bangladeshi people: An online pilot survey early in the outbreak," *J. Affect. Disord.*, vol. 276, pp. 30–37, Nov. 2020, doi: 10.1016/j.jad.2020.06.049.
- [15] A. Sau and I. Bhakta, "Screening of anxiety and depression among seafarers using machine learning technology," *Informatics Med. Unlocked*, vol. 16, p. 100228, 2019, doi: 10.1016/j.imu.2019.100228.
- [16] S. Smys and Jennifer S.Raj, "Analysis of Deep Learning Techniques for Early Detection of Depression on Social Media Network - A Comparative Study," *J. Trends Comput. Sci. Smart Technol.*, vol. 3, no. 1, pp. 24–39, May 2021, doi: 10.36548/jtcsst.2021.1.003.
- [17] M. Tasnim and E. Stroulia, "Detecting Depression from Voice," 2019, pp. 472–478. doi: 10.1007/978-3-030-18305-9_47.
- [18] P. Bobade and M. Vani, "Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Jul. 2020, pp. 51–57. doi: 10.1109/ICIRCA48905.2020.9183244.
- [19] N. Salari *et al.*, "Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis," *Global. Health*, vol. 16, no. 1, p. 57, Dec. 2020, doi: 10.1186/s12992-020-00589-w.
- [20] K. Nemani *et al.*, "Association of Psychiatric Disorders With Mortality Among Patients With COVID-19," *JAMA Psychiatry*, vol. 78, no. 4, p. 380, Apr. 2021, doi: 10.1001/jamapsychiatry.2020.4442.
- [21] J. Šalkevičius, R. Damaševičius, R. Maskeliūnas, and I. Laukienė, "Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals," *Electronics*, vol. 8, no. 9, p. 1039, Sep. 2019, doi: 10.3390/electronics8091039.
- [22] C. Su, Z. Xu, J. Pathak, and F. Wang, "Deep learning in mental health outcome research a scoping review," *Transl. Psychiatry*, vol. 10, no. 1, p. 116, Dec. 2020, doi: 10.1038/s41398-020-0780-3.
- [23] DASS, "Depression Anxiety Stress Scales," 2022. <http://www2.psy.unsw.edu.au/dass/>
- [24] R. E. Shields *et al.*, "Brief mental health disorder screening questionnaires and use with public safety personnel: A review," *Int. J. Environ. Res. Public Health*, vol. 18, no. 7, pp. 1–30, 2021, doi: 10.3390/ijerph18073743.
- [25] R. A. Rahman, K. Omar, S. A. Mohd Noah, M. S. N. M. Danuri, and M. A. Al-Garadi, "Application of Machine Learning Methods in Mental Health Detection: A Systematic Review," *IEEE Access*, vol. 8, pp. 183952–183964, 2020, doi: 10.1109/ACCESS.2020.3029154.