

Review Article

Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges

Jetli Chung and Jason Teo 

Faculty of Computing and Informatics, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

Correspondence should be addressed to Jason Teo; jtwteo@ums.edu.my

Received 5 April 2021; Revised 23 June 2021; Accepted 29 September 2021; Published 5 January 2022

Academic Editor: Aniello Minutolo

Copyright © 2022 Jetli Chung and Jason Teo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The increase of mental health problems and the need for effective medical health care have led to an investigation of machine learning that can be applied in mental health problems. This paper presents a recent systematic review of machine learning approaches in predicting mental health problems. Furthermore, we will discuss the challenges, limitations, and future directions for the application of machine learning in the mental health field. We collect research articles and studies that are related to the machine learning approaches in predicting mental health problems by searching reliable databases. Moreover, we adhere to the PRISMA methodology in conducting this systematic review. We include a total of 30 research articles in this review after the screening and identification processes. Then, we categorize the collected research articles based on the mental health problems such as schizophrenia, bipolar disorder, anxiety and depression, posttraumatic stress disorder, and mental health problems among children. Discussing the findings, we reflect on the challenges and limitations faced by the researchers on machine learning in mental health problems. Additionally, we provide concrete recommendations on the potential future research and development of applying machine learning in the mental health field.

1. Introduction

Mental illness is a health problem that undoubtedly impacts emotions, reasoning, and social interaction of a person. These issues have shown that mental illness gives serious consequences across societies and demands new strategies for prevention and intervention. To accomplish these strategies, early detection of mental health is an essential procedure. Medical predictive analytics will reform the healthcare field broadly as discussed by Miner et al. [1]. Mental illness is usually diagnosed based on the individual self-report that requires questionnaires designed for the detection of the specific patterns of feeling or social interactions [2]. With proper care and treatment, many individuals will hopefully be able to recover from mental illness or emotional disorder [3].

Machine learning is a technique that aims to construct systems that can improve through experience by using advanced statistical and probabilistic techniques. It is

believed to be a significantly useful tool to help in predicting mental health. It is allowing many researchers to acquire important information from the data, provide personalized experiences, and develop automated intelligent systems [4]. The widely used algorithms in the field of machine learning such as support vector machine, random forest, and artificial neural networks have been utilized to forecast and categorize the future events [5].

Supervised learning in machine learning is the most widely applied approach in many types of research, studies, and experiments, especially in predicting illness in the medical field. In supervised learning, the terms, attributes, and values should be reflected in all data instances [6]. More precisely, supervised learning is a classification technique using structured training data [7]. Meanwhile, unsupervised learning does not need supervision to predict. The main goal of unsupervised learning is handling data without supervision. It is very limited for the researchers to apply unsupervised learning methods in the clinical field.

In this paper, the main objective is to provide a systematic literature review, critical review, and summary of the machine learning techniques that are being used to predict, diagnose, and identify mental health problems. Moreover, this paper will propose future avenues for research on this topic. It would also give attention to the challenges and limitations of applying the machine learning techniques in this area. Besides that, potential opportunities and gaps in this field for future research will be discussed. Hence, this paper will contribute to the state of the art in the form of a systematic literature review concerning the machine learning techniques applied in predicting mental health problems. This paper hence contributes a critical summary and potential research directions that could assist researchers to gain knowledge about the methods and applications of big data in the mental health fields.

Although previous papers have been published by reviewing the applications of machine learning approaches toward the mental health field [6, 8], these are general review papers that discuss the applications and concepts of the techniques but do not provide a focused critical summary of the recent gaps in the literature as well as future research directions for this field. As such, this systematic literature review paper aims both to cover recent advancements in this field in addition to providing a focused critical summary concerning the gaps in the literature in terms of the applications of machine learning in the mental health field and to subsequently highlight potential avenues for future research.

The audiences for this paper center around the community of practitioners who are applying machine learning techniques in mental health. Besides that, this paper is targeting the practitioners in the machine learning communities where they can keep updated on the application of machine learning nowadays particularly in the mental health field.

The relevant research papers and documents are gathered and collected through academic publication repositories with specific keywords. Then, the collected documents are identified and categorized into several sections in mental health problems. The performance on the machine learning algorithms or techniques that are used by the researchers is being evaluated by identifying the accuracy, sensitivity, specificity, or area under the ROC curve (AUC).

Hence, the sections of this paper are organized as follows. After Introduction, the Background section presents the information about the mental health prediction problem and, subsequently, machine learning algorithms are discussed. The Methodology section will discuss the strategy of finding the relevant research documents. The Results and Discussion sections will review and examine the machine learning approaches used in predicting mental health problems. Lastly, the Conclusion section will conclude this paper.

2. Background

This review paper follows the standard process of the systematic literature review as shown in Figure 1. First of all, this review paper begins with the planning phase where the

research questions or objectives are investigated and determined. In the planning phase, the data sources are being selected, and then the terms that are related to the topic will be used for searching in the data sources. In conducting the review, several aspects need to be prioritized. For instance, publications of the research articles or papers are identified, the studies of the related topic will be selected, and studies that satisfy the research questions will be chosen. Besides that, the evaluation part will begin by extracting the data from the chosen research articles or papers. Then, further analysis will be carried on the data or evidence from the selected articles and papers. The trends of the research based on the topic will be discussed and investigated. The last part of the process is the discussion and conclusion. The limitations, drawbacks, or gaps of the research will be discussed and examined in this part. Besides that, future directions and potential areas of the research will be investigated and determined. A conclusion will be provided based on the findings from the research.

Figure 2 shows the categorization and classification of the systematic literature review on this topic. The machine learning approaches are being investigated and explored within the scope of mental health problems. The scope of mental health problems is divided into five types of problems, namely, schizophrenia, anxiety and depression, bipolar disorder, posttraumatic stress disorder (PTSD), and mental health problems among children. Additionally, the data of the mental health problems are collected through several domains and sources. This paper will review and highlight the implementation of machine learning models in each mental health problem. Figure 2 presents the machine learning approaches divided into supervised learning, unsupervised learning, ensemble learning, neural networks, and deep learning. Then, the machine learning models are classified based on the type of learning approaches. Besides that, the performances of the machine learning model will be included in this paper to show the efficiency of the machine learning approaches within the mental health field. For instance, the performances such as accuracy, the area under the ROC curve (AUC), *F1*-score, sensitivity, or specificity will be specified and mentioned in this review paper to provide further analysis.

2.1. Mental Health Problems. The World Health Organization (WHO) reports the region-wise status of different barriers in diagnosing mental health problems and encourages researchers to be equipped with the scientific knowledge to address the issue of mental health [9]. Now, there are various techniques to predict the state of mental health due to advancement of technology. Research in the field of mental health has increased recently and contributed to the information and publications about different features of mental health, which can be applied in a wide range of problems [10].

Many steps are involved in diagnosing mental health problems, and it is not a straightforward process that can be done quickly. Generally, the diagnosis will begin with a specific interview that is filled with questions about

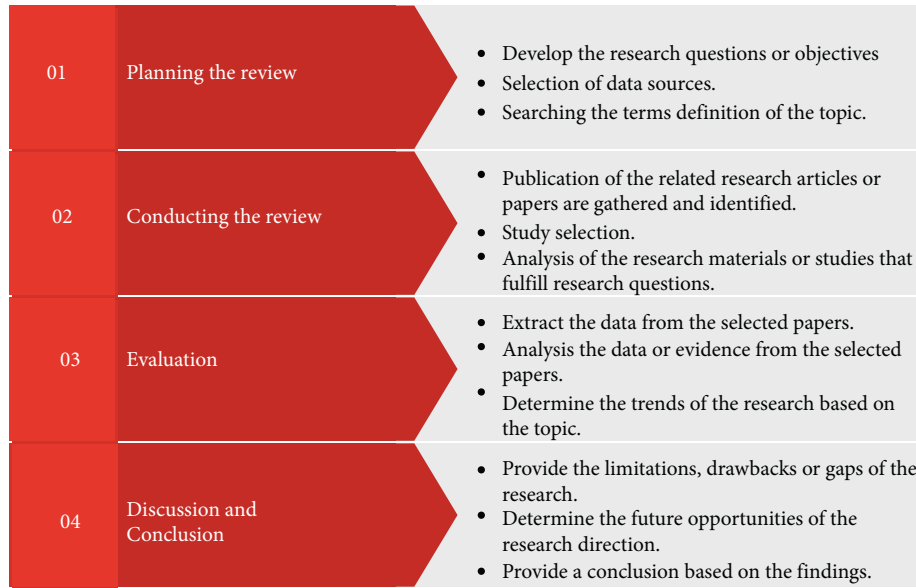


FIGURE 1: Process of the systematic literature review for this topic.

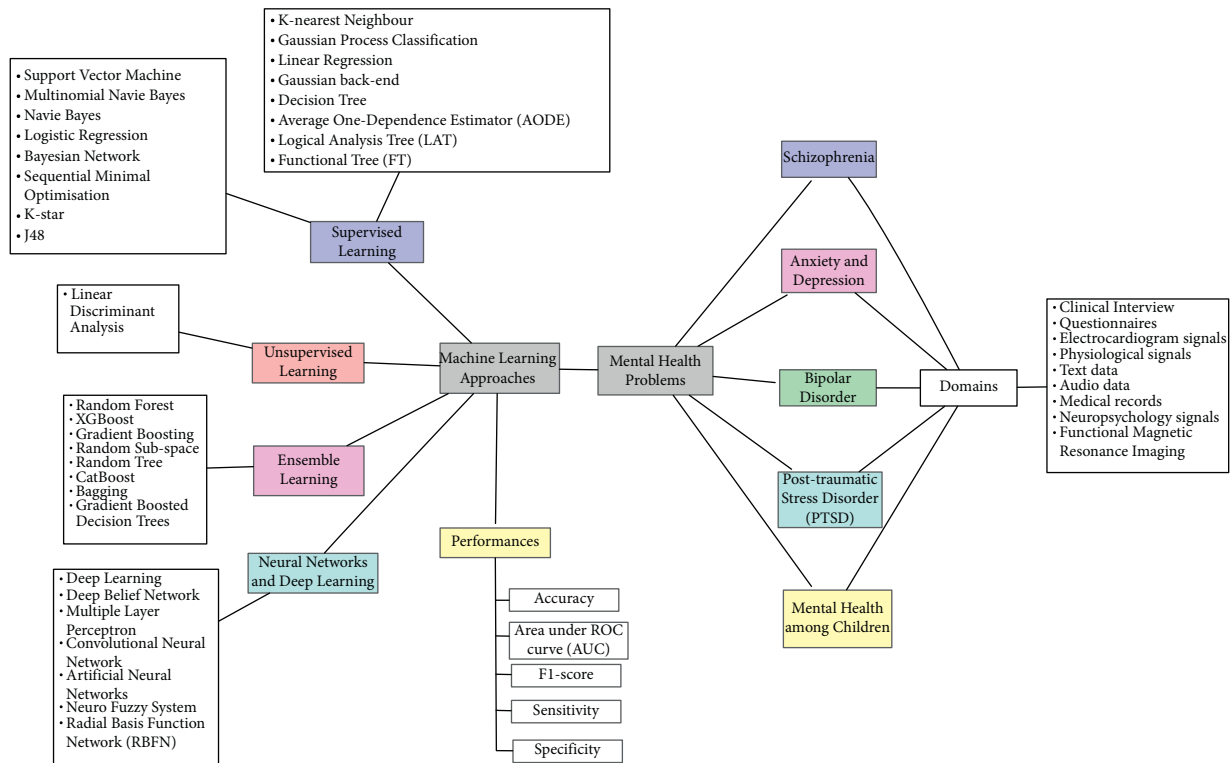


FIGURE 2: Taxonomy of the systematic literature review for this topic.

symptoms, medical history, and physical examination. Besides that, psychological tests and assessment tools are also available and are used to diagnose a person for mental health problems. There are several types of research carried out to investigate and examine the movements of the face to identify certain mental disorders [11].

The increase of research in the mental health field has led to the rise of information in the form of finding suitable

solutions to reduce mental health problems. However, the precise reasons for mental illnesses are still unclear and uncertain.

2.2. Types of Mental Health Problems. Mental illness can affect the cognition, emotion, and behaviour among the people. For children, their ability to learn could be interfered by mental disorders. Besides that, mental illness can cause

inconvenience to the adults, especially in their families, workplaces, and in the society. There are many types of mental disorders commonly known as schizophrenia, depression, bipolar disorder, and anxiety.

Schizophrenia is a mental illness that interrupted by events of psychotic symptoms, which are hallucinations and delusions. Hallucinations are experiences that are not comprehensible to others. Meanwhile, delusions are impressions that are held by the patients although contradicted by the rational and real arguments. Schizophrenia is often diagnosed by symptoms such as social withdrawal, irritability, and increasing strange behaviours. Studies of whether an early diagnosis of such symptoms and intervention could improve the outcomes are still in progress [12].

The primary symptom of depression is an interference of the mood, which is usually severe sadness. Sometimes, anger, irritability, and loss of interests might dominate the symptoms of the depression. In terms of physiological symptoms, sleep disturbance, appetite disturbance, and decreased in energy are commonly shown across cultures. The cognitive symptoms such as slow thinking, suicidal thoughts, and guilt might occur among the patients. Most of the individuals that suffer from depression will have recurrence episodes [13]. Many individuals do not recover completely and they might have a form of chronic mild depression [14].

Bipolar disorder is another mental disorder identified by the episode of mania and depression. Sometimes, there is an episode mixed with both mania and depression. Mania is known by irritability, increased in energy, and decreased need for sleep. Individuals that experience mania often exhibit reckless behaviours. Meanwhile, a depressive episode for bipolar disorder is almost the same as the depression symptoms. Some studies report some recovery to baseline functioning between episodes; however, many patients will have residual symptoms that cause impairment [15].

Another common mental disorder is an anxiety disorder, which is usually identified as an inability to regulate fear or worry. Panic disorders belong to this category, which appears to be unexpected panic attacks and intense fear. The physiological symptoms that are caused by panic disorder include a racing heart, sweating, and dizziness. Generalized anxiety disorder is characterized by excessive worry. Emotional numbness caused by traumatic events characterizes posttraumatic stress disorder (PTSD). Individuals that have a social anxiety disorder are frequently afraid of social situations. Surveys show that delays in seeking professional treatment for an anxiety disorder are widespread [16].

2.3. Data Mining and Machine Learning. In modern days, the management and processing of data have fully grown into a popular topic in the field of computer science. Data mining is knowledge discovery in databases, which is discovering useful patterns and relationships in large volumes of data. Within the medical field, data mining techniques are increasingly applied for tasks such as text expression, drug design, and genomics [17].

Data mining techniques can be separated into two forms, which are supervised learning and unsupervised learning. For unsupervised learning, it determines the object's similarity and detects patterns through the group's data. It can be grouped into clustering, association, summarizing, and sequence discovery [18]. Unsupervised learning is particularly valuable in helping to identify the structure of the data automatically through learning inherent from input data when the data set is unlabelled.

In short, data mining is a crucial technique in the role of computer science. The complexity of the data sets collected can be solved rapidly and swiftly through data mining. In addition, many parties can gain an advantage using data mining for better outcomes and solutions of their challenging problems.

Machine learning is an application of artificial intelligence (AI), which implements systems with the capability to learn and improve from experience without being explicitly programmed. Machine learning has offered essential advantages to a wide range of areas such as speech recognition, computer vision, and natural language processing. It is allowing many researchers to extract meaningful information from the data, provide personalized wisdom, and establish automated intelligent systems [4].

It is believed that machine learning introduced many types of approaches and learning. For instance, the commonly used machine learning approaches are supervised learning and unsupervised learning. Supervised learning is an approach that predicts the outcome result with given labelled data input. Supervised learning is excellent at classification and regression problems. The purpose of this learning is to make sense of data toward the specific measurements. The unsupervised learning is in contrast to the supervised learning, which tries to make sense of data in itself. In unsupervised learning, there are no measurements or guidelines. Additionally, the ensemble learning is a process where the classifiers combined and generated strategically to solve a specific problem. The primary usage of ensemble learning is to improve the performance of a model or reduce the probability of selecting models with poor performance [19]. Moreover, neural networks and deep learning have recently become more well known among machine learning approaches due to their ability to solve many problems such as image recognition, speech recognition, and natural language processing. These approaches are based on the neuronal networks of the brain where they enable the algorithms to learn from the observational data.

In the medical field, machine learning algorithms have been used to discover new drugs, perform radiology analysis, predict epidemic outbreaks, and diagnose diseases. Generally, machine learning algorithms are tools to analyze the massive medical data sets. They are utilized as tools in assisting for medical diagnosis as they became more reliable in their performance.

From time to time, machine learning and data mining approaches continue to develop rapidly. Powerful algorithms and more advanced neural networks, decision trees, gradient boosting, and others were introduced and applied to solve more complicated medical diagnosis problems.

3. Methodology

In this review paper, the planning phase is conducted followed by the searching and analysis phase. Then, the discussion of the relevant documents that are found will be highlighted and summarized in this paper. The conclusions will be presented to conclude this review paper.

Several research questions or objectives for this review paper have been highlighted and investigated. First of all, we want to provide a summary of the latest research on machine learning approaches in predicting mental health problems, which can give useful information to the clinical practice. Besides that, this review paper also will identify the types of machine learning algorithms that have been widely used for this field. We also want to learn and investigate the limitations of the application of machine learning within this field. Moreover, we want to determine the future opportunities or research avenues that can maximize the potential of machine learning approaches within the mental health fields.

For the planning stage, the sources of the database for collecting the research papers and articles are identified. The journals and conferences that are related to the research such as Journal of Psychiatric Research, International Conference on Computational Intelligence and Data Science, and International Conference on Advanced Engineering, Science, Management and Technology have been highlighted in this review paper. Besides that, the reliable publishers such as Springer, ScienceDirect, and IEEE publisher were chosen as the repositories to provide the research papers and articles.

To conduct the searching and analysis, the topic stated has been explored in the following publishers' website. Besides that, the queries such as Machine Learning Algorithms in Mental Health, Psychiatric Medical with Machine Learning Techniques, and Machine Learning in Predicting Mental Health Problems have been used on these sites. The analysis phase is started by finding out and investigating the performance of the machine learning approaches that were used to diagnose or predict mental health problems. Some of the documents and research papers that do not meet the requirement of the topic will be removed.

The discussion phase will begin by reviewing the machine learning algorithms used by the researchers in their experiments to predict the mental problems. Mental health problems will be divided and categorized into several parts. Then, the performance for the machine learning techniques will be described and further analyzed in this phase. Besides that, the research questions will be acknowledged and answered with using the details found during the review of the literature.

The conclusions related to the topic will be highlighted based on the findings and discussion. Moreover, the prediction of the mental health problems by using machine learning approaches will be generalized and summarized.

This review paper will follow the standard PRISMA protocol, which stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses. It is an evidence-based minimum set of items for reporting systematic reviews and meta-analyses. Based on Figure 3, a total of 142 types of

research articles and papers related to this field were found and recorded through the database searching. Besides that, additional records are also identified through other sources. The records that have already been identified will be screened where the duplicate records will be removed or excluded from this review paper. After that, the records with full-text articles that are evaluated for eligibility will be included in this review paper. However, the full-text articles or papers that do not meet the appropriate conditions will be excluded from the review paper for a reason. Hence, a total of 30 research studies related to the topic will be included and highlighted in this paper.

4. Results

In this section, the documents and information related to the machine learning approaches that have been used by the researchers to conduct a prediction or diagnosis for mental health problems will be reviewed and discussed. Moreover, the performance of the machine learning algorithms used will be evaluated and analyzed. The mental health problems will be categorized into several mental health disorders such as schizophrenia, anxiety and depression, bipolar disorder, posttraumatic stress disorder, and children's mental health problems.

A total of 30 research articles were included in this review paper. The research articles were divided and categorized based on the mental health problems such as schizophrenia, bipolar disorder, anxiety and depression, posttraumatic stress disorder, and mental health problems among children. According to Figure 4, six research articles (20.0%) were highlighted in schizophrenia; meanwhile, seven research articles (23.3%) were analyzed in anxiety and depression. Furthermore, there are seven research articles (23.3%) included in bipolar disorder. Eight research articles (26.7%) will be discussed and investigated in posttraumatic stress disorder. There are only two research articles (6.7%) that will be analyzed for mental health problems among children.

The statistics provided in Figure 5 shows the trends of the reviewed research articles and papers based on the years. For instance, the years 2016, 2017, and 2019 show the highest number of papers, which are being included in this review paper. Meanwhile, the lowest number of papers highlighted in this review paper is presented on the years 2010, 2011, 2014, and 2015. Furthermore, this review paper investigated 2 research papers from each year of 2012, 2013, 2018, and 2020.

4.1. Machine Learning Approaches in Predicting Schizophrenia. According to the paper by Greenstein et al., classification of childhood-onset schizophrenia has been performed [20]. The data consist of genetic information, clinical information, and brain magnetic resonance imaging. The authors use a random forest method to calculate the probability of mental disorder. Random forest is being used in this paper because it has lower error rates compared with other methods. The accuracy of 73.7% is obtained after the classification.

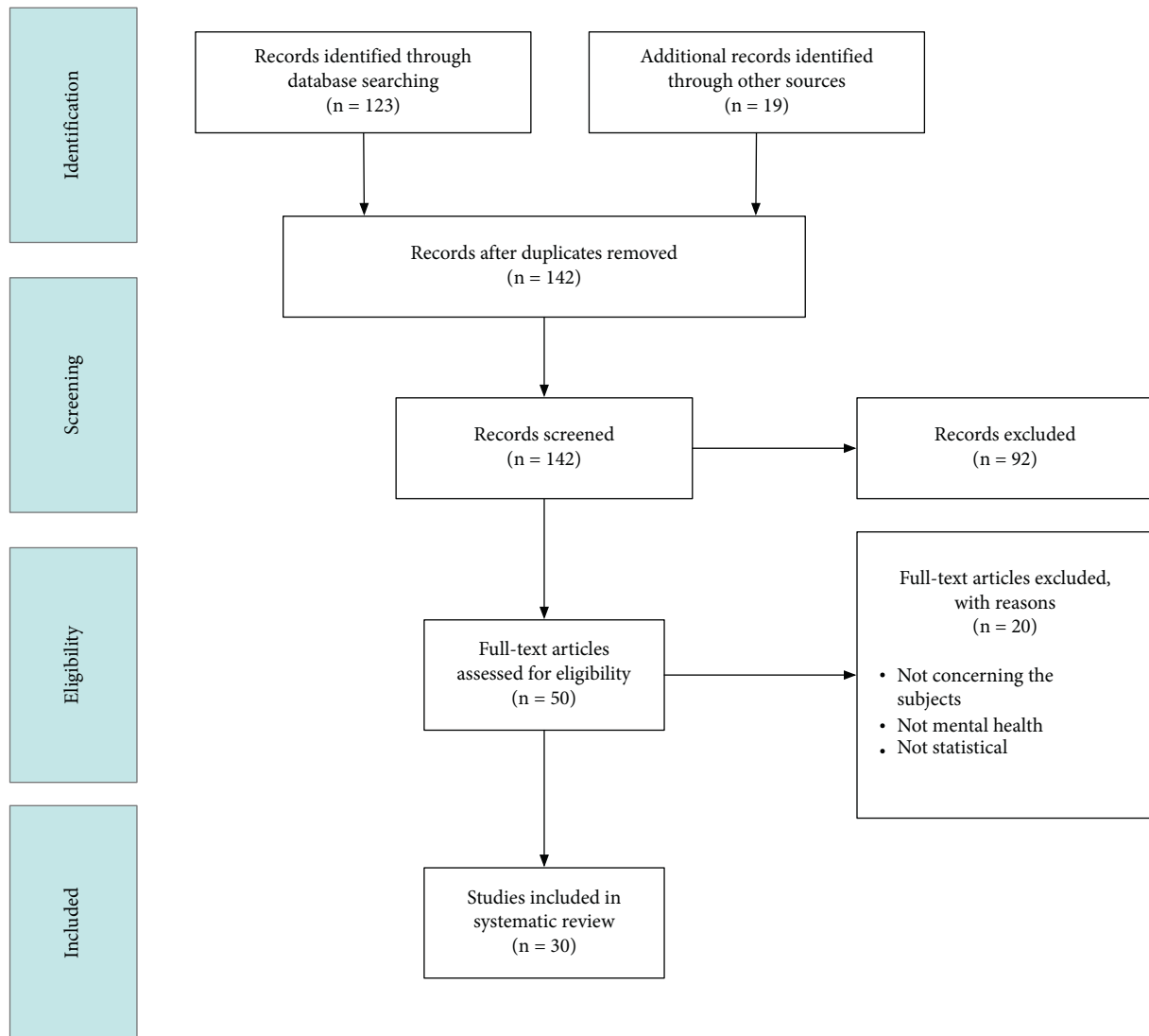


FIGURE 3: PRISMA flow chart of the study selection process.

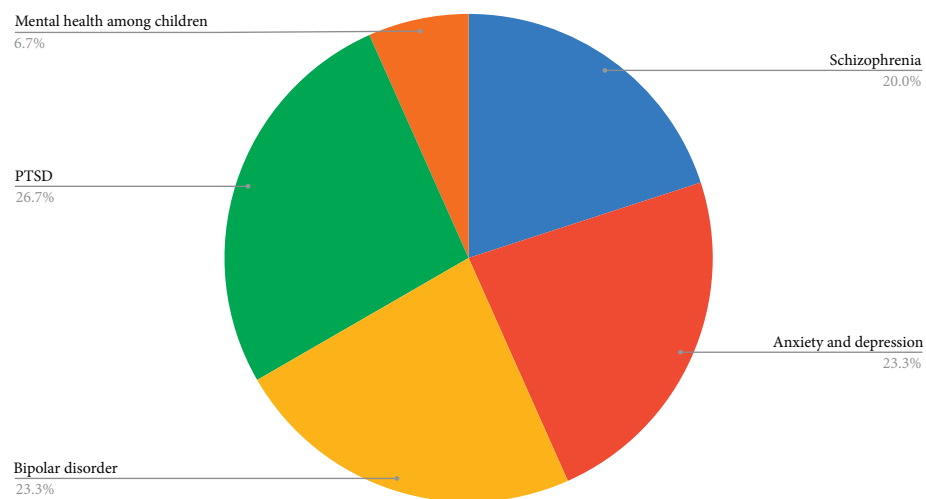


FIGURE 4: Percentage for the number of papers reviewed for each mental health problem.

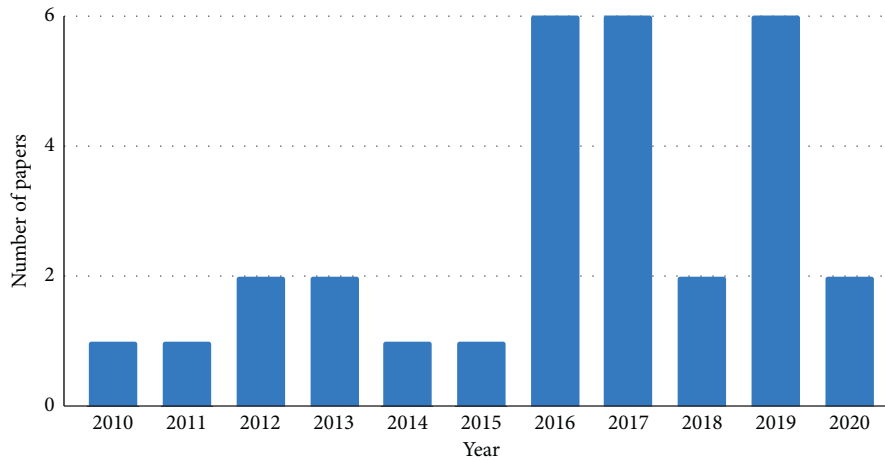


FIGURE 5: Number of papers reviewed per year.

In one of the research works conducted by Jo et al., they used network analysis and machine learning approaches to identify 48 schizophrenia patients and 24 healthy controls [21]. The network properties were rebuilt using the probabilistic brain tractography. After that, machine learning is being applied to label schizophrenia patients and health controls. Based on the result, the highest accuracy is achieved by the random forest model with an accuracy of 68.6% followed by the multinomial naive Bayes with an accuracy of 66.9%. Then, the XGBoost accuracy score is 66.3% and the support vector machine shows an accuracy of 58.2%. Most of the machine learning algorithms show promising levels of performance in predicting schizophrenia patients and healthy controls.

The support vector machine, which is a machine learning model, has been implemented to classify schizophrenia patients [22]. The data set is obtained from the 20 schizophrenia patients and 20 healthy controls. Then, the support vector machine algorithm is used for classification with the help of functional magnetic resonance imaging and single nucleotide polymorphism. After the classification, an accuracy of 0.82 is achieved with the functional magnetic resonance imaging. For the single nucleotide polymorphism, an accuracy of 74% is obtained.

Srinivasagopalan et al. [23] used a deep learning model to diagnose schizophrenia. The National Institute of Health provides the data set for the experiments. The accuracy of each machine learning algorithm is obtained and recorded. The results obtained from the experiment show that deep learning showed the highest accuracy with 94.44%. The random forest recorded an accuracy of 83.33% followed by logistic regression with an accuracy of 82.77%. Then, the support vector machine showed an accuracy of 82.68% in this experiment.

In another study conducted by Pläschke et al., the schizophrenia patients were distinguished from the matched health controls based on the resting-state functional connectivity [24]. Resting-state functional connectivity could be used as a spot of functional dysregulation in specific networks that are affected in schizophrenia. The authors have used support vector machine classification and achieved 68% accuracy.

Pinaya et al. applied the deep belief network to interpret features from neuromorphometry data that consist of 83 healthy controls and 143 schizophrenia patients [25]. The model can achieve an accuracy of 73.6%; meanwhile, the support vector machine obtains an accuracy of 68.1%. The model can detect the massive difference between classes involving cerebrum components. In 2018, Pinaya et al. proposed a practical approach to examine the brain-based disorders that do not require a variety of cases [26]. The authors used a deep autoencoder and can produce different values and patterns of neuroanatomical deviations.

4.2. Machine Learning Approaches in Predicting Depression and Anxiety. A machine learning algorithm is developed to predict the clinical remission from a 12-week course of citalopram [27]. Data are collected from the 1949 patients that experience depression of level 1. A total of 25 variables from the data set are selected to make a better prediction outcome. Then, the gradient boosting method is being deployed for the prediction because of its characteristics that combine the weak predictive models when built. An accuracy of 64.6% is obtained by using the gradient boosting method.

In order to identify depression and anxiety at an early age, a model has been proposed by Ahmed et al. [28]. The model involves psychological testing, and machine learning algorithms such as convolutional neural network, support vector machine, linear discriminant analysis, and K-nearest neighbour have been used to classify the intensity level of the anxiety and depression, which consists of two data sets. Based on the results obtained, the convolutional neural network achieved the highest accuracy of 96% for anxiety and 96.8% for depression. The support vector machine showed a great result and was able to obtain an accuracy of 95% for anxiety and 95.8% for depression. Besides that, the linear discriminant analysis reached the accuracy of 93% for anxiety and 87.9% for depression. Meanwhile, the K-nearest neighbour obtained the lowest accuracy among the models with 70.96% for anxiety and 81.82% for depression. Hence,

the convolutional neural network can be a helpful model to assist psychologists and counsellors for making the treatments efficient.

In the research paper by Sau and Bhakta, they developed a predictive model for diagnosing the anxiety and depression among elderly patients with machine learning technology [29]. Elderly patients have different sociodemographic factors and factors related to health. The data set involved 510 geriatric patients and tested with a tenfold cross-validation method. Then, ten classifiers as shown in Table 1 were selected to predict the anxiety and depression in elderly patients. The metrics of each classifier were evaluated and summarized.

According to Table 1, the highest prediction was obtained by random forest with 89.0%. Then, the J48 accuracy score was 87.8% followed by random subspace with an accuracy of 87.5%. Random tree showed the prediction accuracy with 85.1%; meanwhile, the Bayesian network achieved an accuracy of 79.8%. Next, the naive Bayes and multilayer perceptron achieved the accuracy of 79.6% and 77.8%, respectively. Sequential minimal optimisation and K-star achieved the same accuracy, which is 75.3%. Finally, logistic regression showed the lowest accuracy prediction of 72.4%.

In research conducted by Katsis et al., a system based on physiological signals for the assessment of affective states in anxiety patients has been proposed [30]. The system is proposed to predict the affective state of an individual according to five predefined classes, which are neutral, relaxed, startled, apprehensive, and very apprehensive. The authors use machine learning algorithms in this research such as artificial neural networks, random forest, neuro-fuzzy systems, and support vector machine. The neuro-fuzzy system can obtain the highest accuracy with a score of 84.3% followed by random forest with an accuracy of 80.83%. Meanwhile, the support vector machine and artificial neural networks achieved the accuracies of 78.5% and 77.33%, respectively.

A research paper by Sau and Bhakta shows the prediction of depression and anxiety among seafarers [31]. Seafarers are easily exposed to mental health problems, which typically are depression and anxiety. Hence, machine learning technology has been useful in predicting and diagnosing them for early treatments. The authors were able to obtain a data set of 470 seafarers who were interviewed. In this research conducted by them, features including age, educational qualification, marital status, job profile, type of family, duration of service, existence or nonexistence of heart disease, body mass index, hypertension, and diabetes have been selected to predict the outcome. Five classifiers, which are CatBoost, random forest, logistic regression, naive Bayes, and support vector machine, were chosen on the training data set with 10-fold cross-validation. In order to determine the strength of the machine learning algorithms, the data set with 56 instances are deployed on the trained model. For the training set, the results indicate that the boosting algorithms method CatBoost performs best on this training data set with an accuracy of 82.6%. Random forest has achieved a satisfying accuracy score of 81.2%; meanwhile, logistic regression obtained an accuracy

TABLE 1: Evaluation metrics of ten classifiers in predicting the anxiety and depression among elderly patients.

Classifiers	Evaluation metrics (%)		
	Accuracy	F-measure	AUC
Bayesian network	79.8	79.7	88.9
Naive Bayes	79.6	79.4	85.3
Logistic regression	72.4	72.2	81.1
Multiple layer perceptron	77.8	77.8	85.0
Sequential minimal optimisation	75.3	74.6	75.9
K-star	75.3	75.3	81.4
Random subspace	87.5	87.5	91.7
J48	87.8	87.8	86.0
Random forest	89.0	89.0	94.3
Random tree	85.1	85.1	85.0

score of 77.8%. The support vector machine and naive Bayes obtained 76.1% and 75.8%, respectively. For the test data set, the CatBoost algorithm has performed better than the other machine learning algorithms with a predictive accuracy of 89.3%. Meanwhile, logistic regression has performed very well with the predictive accuracy of 87.5%. Besides, the support vector machine and naive Bayes score the accuracy with the same percentage, which is 82.1%. The random forest shows the lowest accuracy percentage score of 78.6% for the test data set.

Hilbert et al. used machine learning approaches to separate the complicated subjects from healthy ones and distinguish generalized anxiety disorders from major depression without generalized anxiety disorder [32]. For the data set, they used the multimodal behavioural data from a sample of generalized anxiety disorders, healthy persons, and major depression. They applied a binary support vector machine and found out that the prediction of generalized anxiety disorders was difficult when using the clinical questionnaire data. Meanwhile, the input involves the inclusion of cortisol and grey matter volume can reach accuracies of 90.10% and 67.46% for the classification of case and disorder, respectively.

A study has been conducted to detect depression from text and audio by Jerry and others [33]. The study aims to collect the data and improve the analysis from the features of text and voice. The mean of F1-score is analyzed and recorded to determine the best performance among the machine learning algorithms. Tables 2 and 3 show the performance of the machine learning algorithms in detecting depression in text and audio features, respectively.

Based on Tables 2 and 3, random forest has shown the best performance for the text features. With a mean F1-score of 0.73, random forest outperforms all the baseline algorithms. Meanwhile, XGBoost called extreme gradient boosting shows the best performance for audio features with a mean F1-score of 0.50. It is slightly better than the other algorithms.

4.3. Machine Learning Approaches in Predicting Bipolar Disorder. In research performed by Rocha-Rego et al., the authors examined the practicality to determine the bipolar disorder patients from healthy controls by using pattern

TABLE 2: Summary of mean *F1*-score for text features.

Algorithms	Mean <i>F1</i> -score
Gaussian process classification	0.71
Logistic regression	0.69
Neural networks	0.68
Random forest	0.73
Support vector machine	0.72
XGBoost	0.69
K-nearest neighbours	0.67

TABLE 3: Summary of mean *F1*-score for audio features.

Algorithms	Mean <i>F1</i> -score
Gaussian process classification	0.48
Logistic regression	0.48
Neural networks	0.42
Random forest	0.44
Support vector machine	0.40
XGBoost	0.50
K-nearest neighbours	0.49

recognition [34]. The data samples consist of two populations that remitted bipolar disorder patients. A Gaussian process classification algorithm is applied to grey matter and white matter structural magnetic resonance imaging data. The result shows that the accuracy of the algorithm for the grey matters is 73% in study population 1 and 72% in study population 2. Meanwhile, the classification of white matters scored the accuracy of 69% in study population 1 and 78% in population 2.

Grotegerd et al. applied two machine learning models to differentiate depressed bipolar from unipolar patients [35]. The samples involve neuroimaging acquisition where the support vector machine manages to obtain accuracies of 90% in happy against neutral face, 75% in negative against the neutral faces, and 80% when merging the expressions. Meanwhile, the Gaussian process classification shows accuracies of 70% in happy against neutral face, 70% in negative against the neutral faces, and 75% when fusing the expressions.

Valenza et al. suggested a PSYCHE system that functions as some wearable device and the data gathered will be further analyzed for predicting the mood changes in bipolar disorder [36]. The data set consisted of electrocardiogram signals recorded from the patients, and heart rate features from the signals will be selected as the prediction outcome. After applying the support vector machine, an average accuracy of 69% is obtained in predicting the mood states in bipolar disorder.

In another study by Mourão-Miranda et al., the authors applied functional magnetic resonance imaging to explore the differences of the brain activity in patients that have bipolar disorder, major depressive disorder, and healthy controls [37]. The Gaussian process classification algorithm is then trained to determine the bipolar disorder from unipolar depression. The algorithm can achieve an accuracy of 67% with a specificity of 72% and sensitivity of 61%.

In a research article by Roberts et al., a support vector machine is used to distinguish bipolar disorder patients, risk

subjects, and healthy controls [38]. The research involves the data from resting functional connectivity of the left inferior frontal gyrus. The authors used three classes at once to classify the target individual. Based on the result, an overall accuracy of 64.3% was obtained with an independent accuracy of 74.5% in bipolar disorder, 64.5% in risk subjects, and 58.0% in healthy controls.

Another study shows that neuropsychological tests were also applied machine learning techniques published by Akinci et al. [39]. The authors proposed a noninvasive approach to predict bipolar disorder. The different positions of the pupil have been monitored by using the eye pupil detection system. Moreover, the time interval of the pupils when glancing at particular positions and making decisions is managed by the system. With the samples of the data set from the eye pupil, the support vector machine is being applied for the prediction. The prediction accuracy managed to reach an impressive accuracy score of 96.36%.

There are several kinds of research using machine learning approaches and neuropsychological measures to determine the bipolar disorder. Wu et al. conducted an experiment to investigate and determine bipolar disorder among individual patients by using neurocognitive abnormalities [40]. Machine learning known as the LASSO algorithm is then applied to analyze the individual patient with bipolar disorder. The accuracy of 71% and AUC of 0.714 are managed to be obtained through this experiment.

4.4. Machine Learning Approaches in Predicting Posttraumatic Stress Disorder (PTSD). A study conducted by Reece et al. uses the machine learning algorithm random forest to predict the PTSD and depression among the Twitter users [41]. The authors have analyzed more than 243,000 posts from Twitter related to the users that experienced PTSD. Then, the data consisting of PTSD users and healthy controls have been applied in the prediction. With random forest, the authors can predict the PTSD with an AUC score of 0.89.

Leightley et al. applied machine learning techniques for identifying the PTSD among the military forces in the United Kingdom [42]. The authors have collected around 13,690 subjects of the military forces from 2004 to 2009 and used the data as a prediction of PTSD. Various machine learning algorithms are being applied in the prediction. From the experiments, it is found out that random forest has achieved the highest accuracy, which is 97%, in the prediction. Meanwhile, Bagging obtained an accuracy of 95% followed by support vector machine with an accuracy of 91%. The artificial neural network is able to achieve the lowest accuracy among the machine learning algorithms, which is 89%.

Another research about machine learning approaches in PTSD prediction is conducted by Papini et al. [43]. The authors utilized the clinical data, psychological questionnaires, and localization variables when conducting the research. The data set consists of 110 PTSD patients and 231 trauma-exposed controls. A machine learning algorithm known as gradient-boosted decision trees has been built and applied due to its capability in handling the nonlinear

interactions among categorical and continuous features with various distributions. Then, the algorithm was managed to predict the PTSD with an accuracy of 78%.

Besides that, Conrad et al. present the application of machine learning techniques in predicting the PTSD survivors of a civil war in Uganda [44]. The authors use a sample of 441 trauma-exposed subjects as the training data set and 211 trauma-exposed subjects as the new testing data set. Machine learning techniques such as random forest with conditional inference, least absolute shrinkage and selection (LASSO), and logistic regression are being applied to predict the PTSD survivors. Based on the results obtained, using the random forest with conditional inference has shown the highest accuracy of 77.25% compared with the LASSO with an accuracy percentage of 74.88% and logistic regression with an accuracy percentage of 75.36%.

Another research that also used machine learning approaches in the prediction of PTSD from the audio recordings was shown by Marmar et al. [45]. The authors have collected and gathered speech samples from warzone-exposed veterans. Then, the speech attributes that could help in predicting the PTSD such as slower monotonous speech and less change in tonality are extracted from the clinical interviews. Random forest has been used in the prediction and the model can reach the accuracy of 89.1% with AUC of 0.954.

In addition, Vergyri et al. have researched on the audio recordings from the war veterans and compared those with the speech elements of clinicians and patients to predict the PTSD [46]. In the research, they have collected 39 male patients and explored three types of features, which are frame-level features, longer-range prosodic features, and lexical features. Then, they selected Gaussian backend, decision tree, neural network classifiers, and boosting for the prediction model. Using several machine learning models, an overall accuracy of 77% can be generated in the prediction of PTSD.

Based on the study conducted by Salminen et al., the authors have applied a support vector machine in diagnosing PTSD among war veterans by using cortical and subcortical imaging [47]. The data collected are from 97 war veterans who are exposed to the early stress life and participate in the military encounters. Furthermore, they selected the surface in the right posterior cingulate as a major attribute in the classification. The authors can obtain the diagnosis of PTSD with a low accuracy of 69%.

Additionally, Rangaprakash et al. have introduced support vector machines in identifying areas related to the PTSD by combining the functional magnetic resonance imaging and diffusion tensor information [48]. A sample of 87 male soldiers was collected and analyzed to obtain related information and features. After the classification by using the support vector machine, the authors achieved the accuracy percentage of 83.59% and found a relationship between hippocampal-striatal hyperconnectivity and PTSD.

4.5. Machine Learning Approaches in Predicting Mental Health Problems among Children. In the research paper by Sumathi and Poorna, the authors have predicted mental

health problems among children by various machine learning approaches [49]. The factors, symptoms, and psychological tests of the mental health problems are being observed by professionals. The data set is obtained from a clinical psychologist containing 60 instances. Several features and attributes have been selected for the classification process. Different machine learning algorithms were applied to this problem to test their prediction accuracies.

From the result shown in Table 4, the machine learning technique called average one-dependence estimator (AODE) has recorded 71% of accuracy. Meanwhile, MLP shows the highest accuracy, which is 78%. Next is logical analysis tree (LAT) with 70% accuracy; meanwhile, the multiclass classifier is at 58% of accuracy. Another machine learning technique called radial basis function network (RBFN) records the accuracy with 57%. K-star and functional tree (FT) obtained the same accuracy score of 42% in this experiment.

Based on the recent research conducted by Tate et al., the authors applied machine learning algorithms to predict the mental health problems among children [50]. The data consist of a total of 7638 twins from the Child and Adolescent Twin Study in Sweden. They used 474 predictors that are extracted from the register data and parental data. Then, the Strengths and Difficulties Questionnaire was applied to determine the outcome. Based on the result from the test set, the random forest showed the highest AUC of 0.739 followed by the support vector machine with the AUC of 0.736. The neural network recorded the AUC with a score of 0.705. Then, the logistic regression scored an AUC of 0.700, and the XGBoost performed on the test set with an AUC of 0.692.

4.6. Summary. The articles utilizing machine learning approaches in predicting mental health problems have been listed in Table 5.

5. Critical Analysis and Discussion

In this paper, there are a total of 30 research papers that have been reviewed and evaluated in which the use of machine learning techniques or approaches in predicting mental health problems is highlighted. The research papers and articles have been divided and categorized into different types of mental health problems such as schizophrenia, depression, anxiety, bipolar disorder, and PTSD. Besides that, the performance of the machine learning mechanisms that are being applied has been highlighted because it could provide benefits within the medical field in data mining or big data fields.

Based on the summary table provided in Table 5, there are 6 articles that applied various machine learning approaches to identify and predict the schizophrenia patients with different data sets [20–25]. Several research projects have been conducted to analyze and classify depression and anxiety. 7 research papers have been reviewed in this paper to evaluate the performance of machine learning techniques in determining the depression and anxiety among people [27–33]. Besides that, there are 7 studies on the mental

TABLE 4: Accuracy of machine learning techniques in predicting the mental health problems among children.

Machine learning	Accuracy (%)
Average one-dependence estimator (AODE)	71
Multilayer perceptron	78
Logical analysis tree (LAT)	70
Multiclass classifier	58
Radial basis function network (RBFN)	57
K-star	42
Functional tree (FT)	42

health problem bipolar disorder. The studies are being conducted to predict the bipolar disorder among patients by using the machine learning approaches [34–40]. In addition, research on the application of the machine learning in predicting the PTSD has been gaining popularity, and thus there are 8 research articles highlighted this problem in this paper [41–48]. There are also 2 articles that predict the mental health problems among children with various machine learning approaches [49, 50].

In terms of sample data sets used by the researchers, the data sets used for the classification are mostly small size, which is below 100 subjects. For example, the authors Jo et al. [21], Yang et al. [22], Rocha-Rego et al. [34], Grotegerd et al. [35], Mourão-Miranda et al. [37], Akinci et al. [39], Wu et al. [40], Vergyri et al. [46], Salminen et al. [47], and Rangaprakash et al. [48] have applied small size of sample data for the classifications. Moreover, some studies are conducted by using a partial large size of the data set, which is above 100 subjects. Some of the research papers highlighted such amount of the data set, which are Greenstein et al. [20], Srinivasagopalan et al. [23], Pläschke et al. [24], Pinaya et al. [25], Roberts et al. [38], Reece et al. [41], and Marmar et al. [45]. The researchers also performed the prediction with large size of data set. For instance, Chekroud et al. [27], Sau and Bhakta [29], Sau and Bhakta [31], Leightley et al. [42], Papini et al. [43], Conrad et al. [44], and Tate et al. [50] have utilized large size of data set in the prediction of the mental health problems, which are above 300 subjects. Not only that, some authors such as Ahmed et al. [28], Katsis et al. [30], Hilbert et al. [32], Xu et al. [33], Valenza et al. [36], and Sumathi and Poorna [49] have conducted the classification experiments with different types of data set. In the research articles, the data sets consisting of interviews, questionnaires, electrocardiogram signals, physiological signals, and text and audio data have been applied to perform the classifications.

According to the research papers provided, the experiments of the classification are conducted with various machine learning models. It is undeniable that machine learning models such as random forest and support vector machine have been the most popular choice to be applied in the experiments. This is because random forest and support vector machine at most of the time are able to provide an excellent performance in terms of the accuracy. For example, Greenstein et al. [20], Jo et al. [21], Yang et al. [22], Srinivasagopalan et al. [23], Pläschke et al. [24], Pinaya et al. [25], Sau and Bhakta [29], Ahmed et al. [28], Katsis et al.

[30], Sau and Bhakta [31], Hilbert et al. [32], Xu et al. [33], Grotegerd et al. [35], Valenza et al. [36], Roberts et al. [38], Akinci et al. [39], Reece et al. [41], Leightley et al. [42], Conrad et al. [44], Marmar et al. [45], Salminen et al. [47], Rangaprakash et al. [48], and Tate et al. [50] have applied the random forest and support vector machine in the classification of the mental health problems.

According to the results provided by the authors, they usually present the accuracy as the performance measurement level for the machine learning models in predicting the mental health problems. Hence, this paper will highlight the performance of machine learning used in the experiments for each mental health problem that are stated.

First of all, the support vector machine shows an unsatisfying performance in classifying the schizophrenia patients where the accuracy is lower than 70% as stated by Jo et al. [21], Pläschke et al. [24], and Pinaya et al. [25]. However, the support vector machine presents an excellent accuracy as stated by Yang et al. [22] and Srinivasagopalan et al. [23]. Moreover, random forest has provided great accuracy in the experiments conducted by Greenstein et al. [20] and Srinivasagopalan et al. [23], but Jo et al. [21] show that random forest obtains a low accuracy, which is 68.9%. A research article published by Srinivasagopalan et al. [23] shows deep learning can provide an excellent accuracy, which is 94.44%, in classifying the schizophrenia problem.

In classifying the depression and anxiety cases with machine learning models, the research shows a better result in terms of accuracy for the studies conducted. Most of the research articles show that machine learning models have obtained the accuracy of above 70%. However, Chekroud et al. [27] present that gradient boosting achieves the accuracy of 64.6%. Meanwhile, the convolutional neural network has obtained excellent performance with an accuracy of 96.0% for anxiety classification and 96.8% for the depression classification as stated in the article by Ahmed et al. [28]. Besides, the random forest and support vector machine perform very well in classifying the depression and anxiety cases as stated in the research articles by Sau and Bhakta [29], Katsis et al. [30], Sau and Bhakta [31], and Hilbert et al. [32].

Other research articles show different results obtained from bipolar disorder prediction with machine learning models. Rocha-Rego et al. [34] and Grotegerd et al. [35] can apply a machine learning model known as Gaussian process classification and obtain average performance, which is above 70%. Meanwhile, Mourão-Miranda et al. [37] obtained an accuracy of 67% by using the Gaussian process classification. Besides that, the support vector machine provides an unsatisfying performance with an accuracy of 64.3% in Roberts et al. [38] and an accuracy of 69% in Valenza et al. [36]. However, this machine learning model can reach an accuracy score of 96.36% when predicting bipolar disorders as stated by Akinci et al. [39].

When predicting mental health problems for PTSD, machine learning models commonly used are random forest and support vector machine. In the reviewed research articles, random forest has shown an excellent performance in predicting PTSD individuals. For example, Leightley et al.

TABLE 5: Summary of the machine learning approaches within mental health problems.

Author, year	Mental health problem	Sample data set	Machine learning model	Performances	Comments
Greenstein, 2012 [20]	Schizophrenia	(i) 98 childhood-onset schizophrenia (ii) 99 healthy controls	Random forest	Accuracy: 73.7%	Regional brain measured have been chosen to provide lower resolution compared to the higher resolution voxel-wise measures
Jo et al., 2020 [21]	Schizophrenia	(i) 48 schizophrenia patients (ii) 24 healthy controls	(i) Random forest (ii) Multinomial naive Bayes (iii) XGBoost (iv) Support vector machine	Accuracy: (i) Random forest: 68.9% (ii) Multinomial naive Bayes: 66.9% (iii) XGBoost: 66.3% (iv) Support vector machine: 58.2%	
Yang et al., 2010 [22]	Schizophrenia	(i) 20 schizophrenia patients (ii) 20 healthy controls	Support vector machine	Accuracy: (i) 0.82 with functional magnetic resonance imaging (ii) 0.74 with single nucleotide polymorphism	
Srinivasagopalan et al., 2019 [23]	Schizophrenia	(i) 69 schizophrenia patients (ii) 75 controls	(i) Deep learning (ii) Support vector machine (iii) Random forest (iv) Logistic regression	Accuracy: (i) Deep learning: 94.44% (ii) Support vector machine: 82.68% (iii) Random forest: 83.33% (iv) Logistic regression: 82.77%	Young-old classification was dependent on all networks and outperformed clinical classification
Pläschke et al., 2017 [24]	Schizophrenia	(i) 86 schizophrenia patients (ii) 84 healthy controls	Support vector machine	Accuracy: 68%	
Pinaya et al., 2016 [25]	Schizophrenia	(i) 143 schizophrenia patients (ii) 83 healthy controls	(i) Deep belief network (ii) Support vector machine	Accuracy: (i) Deep belief network: 73.6% (ii) Support vector machine: 68.1%	
Chekroud et al., 2016 [27]	Depression	1949 patients with level 1 of depression	Gradient boosting	Accuracy: 64.6%	The random forest model has been tested with another data set and achieved the accuracy of 91.0%
Sau and Bhakta, 2017 [29]	Depression and anxiety	510 elderly patients	(i) Bayesian network (ii) Naive Bayes (iii) Logistic regression (iv) Multilayer perceptron (v) Sequential minimal optimisation (vi) K-star (vii) Random subspace (viii) J48 (ix) Random forest (x) Random tree	Accuracy: (i) Bayesian network: 79.8% (ii) Naive Bayes: 79.6% (iii) Logistic Regression: 72.4% (iv) Multilayer perceptron: 77.8% (v) Sequential minimal optimisation: 75.3% (vi) K-star: 75.3% (vii) Random subspace: 87.5% (viii) J48: 87.8% (ix) Random forest: 89.0% (x) Random tree: 85.1%	

TABLE 5: Continued.

Author, year	Mental health problem	Sample data set	Machine learning model	Performances	Comments
Ahmed et al., 2019 [28]	Depression and anxiety	Data set of depression and anxiety	(i) Convolutional neural network (ii) Support vector machine (iii) Linear discriminant analysis (iv) K-nearest neighbour	Highest accuracy achieved by convolutional neural network: 96.0% for anxiety and 96.8% for depression	
Katsis et al., 2011 [30]	Anxiety	Physiological signals among anxiety patients	(i) Artificial neuro networks (ii) Random forest (iii) Neuro-fuzzy system (iv) Support vector machine	Accuracy: (i) Artificial neuro networks: 77.3% (ii) Random forest: 80.83% (iii) Neuro-fuzzy system: 84.3% (iv) Support vector machine: 78.5%	Overall classification accuracy is 84.3%
Sau and Bhakta, 2019 [31]	Depression and anxiety	Data set of 470 seafarers	(i) CatBoost (ii) Logistic regression (iii) Support vector machine (iv) Naive Bayes (v) Random forest	Accuracy: (i) CatBoost: 89.3% (ii) Logistic regression: 87.5% (iii) Support vector machine: 87.5% (iv) Naive Bayes: 82.1% (v) Random forest: 78.6%	CatBoost has achieved the highest accuracy of 89.3% and highest precision of 89.0%
Hilbert et al., 2017 [32]	Anxiety	Multimodal behavioural data with sample of anxiety disorders, healthy persons and major depression	Support vector machine	Accuracy: (i) 90.10% for the case classification (ii) 67.46% for the disorder classification	

TABLE 5: Continued.

Author, year	Mental health problem	Sample data set	Machine learning model	Performances	Comments
Jerry et al., 2019 [33]	Depression	Text and audio data sets	(i) Gaussian process classification (ii) Logistic regression (iii) Neural networks (iv) Random forest (v) Support vector machine (vi) XGBoost (vii) K-nearest neighbours	Mean F1-score for text data set: (i) Gaussian process classification: 0.71 (ii) Logistic regression: 0.69 (iii) Neural networks: 0.68 (iv) Random forest: 0.73 (v) Support vector machine: 0.72 (vi) XGBoost: 0.69 (vii) K-nearest neighbours: 0.67 Mean F1-score for audio data set: (i) Gaussian process classification: 0.48 (ii) Logistic regression: 0.48 (iii) Neural networks: 0.42 (iv) Random forest: 0.44 (v) Support vector machine: 0.40 (vi) XGBoost: 0.50 (vii) K-nearest neighbours: 0.49	For the text data set, random forests show the best performance. For the audio data set, XGBoost show the best performance.
Rocha-Rego et al., 2014 [34]	Bipolar disorder	(i) 40 subjects with bipolar disorder (ii) 40 subjects of healthy controls	Gaussian process classification	Accuracy: 69–78% Sensitivity: 64–77% Specificity: 69–99%	
Grotegerd et al., 2013 [35]	Bipolar disorder	(i) 10 subjects with bipolar disorder (ii) 10 subjects with unipolar disorder (iii) 10 healthy controls	(i) Gaussian process classification (ii) Support vector machine	Accuracy: (i) Gaussian process classification: 70% (ii) Support vector machine: 70%	
Valenza et al., 2016 [36]	Bipolar disorder	Electrocardiogram signals from the patients	Support vector machine	Accuracy: 69%	
Mourão-Miranda et al., 2012 [37]	Bipolar disorder	(i) 18 subjects with bipolar disorder (ii) 18 subjects with unipolar disorder (iii) 18 healthy controls	Gaussian process classification	Accuracy: 67% Specificity: 72% Sensitivity: 61%	
Roberts et al., 2016 [38]	Bipolar disorder	(i) 49 bipolar disorder patients (ii) 71 at-risk subjects (iii) 80 healthy controls	Multiclass support vector machine	Overall accuracy: 64.3%	
Akinci et al., 2012 [39]	Bipolar disorder	(i) 40 subjects with bipolar disorder (ii) 55 healthy controls	Support vector machine	Accuracy: 96.36%	
Wu et al., 2016 [40]	Bipolar disorder	(i) 21 subjects of bipolar disorder (ii) 21 healthy controls	LASSO	Accuracy: 71% AUC: 0.714	
Reece et al., 2017 [41]	PTSD	(i) 63 PTSD (ii) 111 healthy controls	Random forest	AUC: 0.89	There are two categories for detecting the PTSD and depression

TABLE 5: Continued.

Author, year	Mental health problem	Sample data set	Machine learning model	Performances	Comments
Leightley et al., 2018 [42]	PTSD	13,690 subjects of the military forces from 2004 to 2009	(i) Support vector machine (ii) Random Forest (iii) Artificial neural networks (iv) Bagging	Accuracy: (i) Support vector machine: 91% (ii) Random forest: 97% (iii) Artificial neural networks: 89% (iv) Bagging: 95% Accuracy: 78%	Exploitation of alcohol, gender, and deployment status are the variables affected to the performance
Papini et al., 2018 [43]	PTSD	(i) 110 PTSD patients (ii) 231 trauma-exposed controls	Gradient-boosted decision trees	AUC: 0.85 Sensitivity: 69% Specificity: 83% Accuracy:	
Conrad et al., 2017 [44]	PTSD	(i) 441 trauma-exposed subjects as training data set (ii) 211 trauma-exposed subjects as testing data set	(i) Random forest with conditional inference (ii) LASSO (iii) Linear regression	(i) Random forest with conditional inference: 77.25% (ii) LASSO: 74.88% (iii) Linear regression: 75.36%	
Marmar et al., 2019 [45]	PTSD	(i) 52 subjects of PTSD (ii) 77 trauma-exposed controls	Random forest	Accuracy: 89.1% AUC: 0.954	
Vergyri et al., 2015 [46]	PTSD	(i) 15 subjects of PTSD (ii) 24 trauma-exposed controls	(i) Gaussian backend (ii) Decision tree (iii) Neural network (iv) Boosting	Overall accuracy: 77%	Speech features have particular power for the prediction of the PTSD
Salminen et al., 2019 [47]	PTSD	97 war veterans	Support vector machine	Accuracy: 69% Sensitivity: 58% Specificity: 81%	The important feature chosen is the surface area in the right posterior cingulate
Rangaprakash et al., 2017 [48]	PTSD	87 male soldiers	Support vector machine	Accuracy: 83.59%	PTSD is associated with hippocampal-striatal hyperconnectivity
Sumathi and Poorna, 2016 [49]	Mental health problems among children	Data set from interviews which contained 60 instances	(i) Average one-dependence estimator (AODE) (ii) Multilayer perceptron (iii) Logical analysis tree (LAT) (iv) Radial basis function network (RBFN) (v) K-star (vi) Functional tree (FT)	Accuracy: (i) AODE: 71% (ii) Multilayer perceptron: 78% (iii) LAT: 70% (iv) RBFN: 57% (v) K-star: 42% (vi) FT: 42%	The highest accuracy is achieved by multilayer perceptron, which is 78%
Tate et al., 2020 [50]	Mental health problems among children	7638 twins from the Child and Adolescent Twin Study in Sweden	(i) Random forest (ii) Support vector machine (iii) Neural network (iv) Logistic regression (v) XGBoost	AUC: (i) Random forest: 0.739 (ii) Support vector machine: 0.736 (iii) Neural network: 0.705 (iv) Logistic regression: 0.700 (v) XGBoost: 0.692	The performances of machine learning models are close to each other

[42] have achieved a percentage of 97% of accuracy with random forest. In addition, Marmar et al. [45] and Reece et al. [41] applied random forest in their studies to predict PTSD individuals. From the results, Marmar et al. [45] obtained an accuracy score of 89.1% with random forest; meanwhile, Reece et al. [41] managed to reach an AUC of 0.89 with random forest. In a research article by Leightley et al. [42], the authors have utilized the support vector machine and obtained a satisfying accuracy of 91%. Besides that, Rangaprakash et al. [48] have shown that the support vector machine can achieve an accuracy of 83.59% when classifying the PTSD among male soldiers. However, the support vector machine shows some drawbacks when predicting the PTSD among war veterans in Salminen et al. [47] where it only obtained an accuracy of 69%.

Based on the research articles published by Sumathi and Poorna [49] and Tate et al. [50], the authors have used machine learning models when predicting the mental health problems among the children. From the obtained results, Sumathi and Poorna showed that multilayer perceptrons can achieve an accuracy of 78%, which is the highest accuracy among machine learning models [49]. Moreover, Tate et al. applied machine learning models to predict the mental health problems with twins children data set. They obtained the highest AUC by using random forest, which is 0.739, followed by support vector machine, which is 0.736 [50].

5.1. Gaps in the Literature. In this section, it would be crucial to provide the challenges and limitations encountered by the researchers to learn the gaps in the literature of machine learning approaches in this field.

5.1.1. Small Sample Size. It is notable that most of the reviewed research lack a sample size or applying a small sample size in their experiments. Even though machine learning can exhibit robustness when analyzing the large sample size, certain approaches can perform with a small sample without compromising the accuracy depending on the settings toward the model applied in the experiments. Vabalas et al. mentioned that usage of the small sample is common in the field of mental health because of the cost that is related to the data collection that involves the human participants and the experimental rules with different conditions are still under development [51].

5.1.2. Insufficient Validation. Due to the small sample sizes and insufficient acceptable validation from external sources, many types of research are still in a proof-of-concept stage. For example, structural neuroimaging research projects are usually carried out in subjects who already had mental health illness. This is difficult to decide whether structural brain alterations are the risk factors, result, or illness source. The researchers should cooperate with a clinical professional to provide important information such as validation, truth, and biases, which could lead to the analysis of data, improve accuracy, and manage deployment risks [52].

5.1.3. Limited Exploration in Deep Learning. It is believed that deep learning algorithms have been successful for a few applications, especially in healthcare domains. However, there is still limited exploration in the use of deep learning algorithms for mental health. Besides that, deep learning algorithms are treated as a black box, leading the researchers to have a challenging time trying to explain why and how these deep learning algorithms work. Recent studies show there are attempts to open the black box of deep learning algorithms [53]. Such exploration is very crucial for the researchers to convince the medical professionals to apply the predictive mental health system.

5.1.4. Lack of Real-Life Testing. Although machine learning can show the researchers about the prediction on mental health, there is still a lack of testing being applied in real life due to several reasons. Many medical professionals still doubt the accuracy of automated methods such as machine learning, as well as issues of consistency and difficulty when applying the machine learning predictive systems to real-world medical practices. Dang et al. stated that there is no standard way to collect high-quality data, difficulty in achieving the labels, which cause the supervised learning approaches to be inconsistent, and also the lack of acknowledging the best practices in handling machine learning models [54]. Such challenges and reasons could reduce the real-life application of machine learning models in the mental health field.

5.2. Avenues for Future Research. Next, this paper will highlight the specific approaches that could help in the development of the research toward the effectiveness of the machine learning application in this field.

5.2.1. Exploration in Deep Learning. The success of applying machine learning approaches in mental health prediction can be expanded to include deep learning approaches. Such approaches could even predict mental health problems together with diagnosis of other chronic diseases such as cancer, diabetes, and others. Architectures of deep learning in processing the image can be useful to identify and predict mental health problems from facial expression. In this context, deep learning architectures hopefully could be combined with memory [55] and attention mechanisms [56] to build greater accuracy clinical architectures.

5.2.2. High-Quality Data. In order to develop more accurate predictive tools, data such as sociodemographics, speech, medical report profiles, and facial expressions of the patients can be recorded or taken via photography combined with magnetic resonance imaging of the brain. In this approach, the data have a larger volume where the deep learning algorithms can be useful and applied. Obtaining such a detailed and large data set shows a challenge for the mental health field and requires immediate collaboration among institutes and organizations [57].

5.2.3. Accurate Predictive Tools. The application of new models to predict the clinical results should be given the research opportunity. Besides, web-based predictors and medical analytics tools should be developed to transform the effective predictive models into useful clinical decision systems such as for identifying the different types of mental disorders, medication plans, as well as preventive plans. For instance, Psycho Web is being developed where the application allows users to collect and predict the data from mental health patients using machine learning [58]. However, this application is still in its infancy and undergoing continual improvements.

5.2.4. Explainable Model. Performance of machine learning models and being explainable are necessary for mental health problems. Medical professionals need to understand the underlying system of prediction and classification very well before practising it in the real world and with patients. Making the results obtained by these models understandable should be the main priority toward establishing reliable systems. In a paper conducted by Holzinger et al., the authors encouraged an innovative and interactive explainable approach called counterfactual graphs for the beneficial future interaction between humans and artificial intelligence [59].

5.2.5. Transfer Learning and Flexible Algorithms. Transfer learning is an algorithm developed for adaptability to different purposes where it could help to improve the generalization performance of machine learning models. Transfer learning has been widely applied in fields that require image analysis, which could be useful to incorporate in clinical settings [60]. Meanwhile, flexible algorithms will become the main challenge toward mental health because of heterogeneity in the input data. Machine learning models need to have a life-long framework as it can help preventing the catastrophic forgetting [61]. Achieving the best results with these future opportunities will need cooperative efforts between the data researchers, computer scientists, and medical professionals.

6. Conclusion

Many different techniques and algorithms had been introduced and proposed to test and solve the mental health problems. There are still many solutions that can be refined. In addition, there are still many problems to be discovered and tested using a wide variety of settings in machine learning for the mental health domain. As classifying the mental health data is generally a very challenging problem, the features used in the machine learning algorithms will significantly affect the performance of the classification.

The existing studies and research show that machine learning can be a useful tool in helping understand psychiatric disorders. Besides that, it may also help distinguish and classify the mental health problems among patients for further treatment. Newer approaches that use data that arise from the integration of various sensor modalities present in

technologically advanced devices have proven to be a convenient resource to recognize the mood state and responses from patients among others.

It is noticeable that most of the research and studies are still struggling to validate the results because of insufficiency of acceptable validated evidence, especially from the external sources. Besides that, most of the machine learning might not have the same performance across all the problems. The performance of the machine learning models will vary depending on the data samples obtained and the features of the data. Moreover, machine learning models can also be affected by preprocessing activities such as data cleaning and parameter tuning in order to achieve optimal results.

Hence, it is very important for researchers to investigate and analyze the data with various machine learning algorithms to choose the highest accuracy among the machine learning algorithms [62]. Not only that, challenges and limitations faced by the researchers need to be managed with proper care to achieve satisfactory results that could improve the clinical practice and decision-making.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The corresponding author was supported by a research grant from the Ministry of Higher Education, Malaysia (Fundamental Research Grant Scheme (FRGS), Dana Penyelidikan, Kementerian Pengajian Tinggi, FRGS/1/2019/ICT02/UMS/01/1). The APC was funded by Universiti Malaysia Sabah.

References

- [1] G. Miner, L. Miner, M. Goldstein et al., *Practical Predictive Analytics and Decisioning Systems for Medicine: Informatics Accuracy and Cost-Effectiveness for Healthcare Administration and Delivery Including Medical Research*, Academic Press, Inc., Cambridge, MA, USA, 1st edition, 2016.
- [2] M. Hamilton, "Development of a rating scale for primary depressive illness," *British Journal of Social and Clinical Psychology*, vol. 6, no. 4, pp. 278–296, 1967.
- [3] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, Reston, VA, USA, 2013.
- [4] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [5] F. Dabek and J. J. Caban, "A neural network based model for predicting psychological conditions," in *Brain Informatics and Health*, pp. 252–261, Springer International Publishing, Berlin, Germany, 2015.
- [6] G. Cho, J. Yim, Y. Choi, J. Ko, and S.-H. Lee, "Review of machine learning algorithms for diagnosing mental illness," *Psychiatry Investigation*, vol. 16, no. 4, pp. 262–269, 2019.

- [7] E. G. Pintelas, T. Kotsilieris, I. E. Livieris, and P. Pintelas, "A review of machine learning prediction methods for anxiety disorders," in *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion—DSAI 2018*, Thessaloniki, Greece, 2018.
- [8] R. B. Rutledge, A. M. Chekroud, and Q. J. Huys, "Machine learning and big data in psychiatry: toward clinical applications," *Current Opinion in Neurobiology*, vol. 55, pp. 152–159, 2019.
- [9] 130 Executive Board, "Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social workers at the country level," 2012, <https://apps.who.int/iris/handle/10665/23741>.
- [10] M. Hadzic, M. Chen, and T. S. Dillon, "Towards the mental health ontology," in *Proceedings of the 2008 IEEE International Conference on Bioinformatics and Biomedicine*, Philadelphia, PA, USA, 2008.
- [11] J. Diederich, A. Al-Ajmi, and P. Yellowlees, "Ex-ray: data mining and mental health," *Applied Soft Computing*, vol. 7, no. 3, pp. 923–928, 2007.
- [12] P. D. McGorry, A. R. Yung, L. J. Phillips et al., "Randomized controlled trial of interventions designed to reduce the risk of progression to first-episode psychosis in a clinical sample with subthreshold symptoms," *Archives of General Psychiatry*, vol. 59, no. 10, pp. 921–928, 2002.
- [13] T. I. Mueller, A. C. Leon, M. B. Keller et al., "Recurrence after recovery from major depressive disorder during 15 years of observational follow-up," *American Journal of Psychiatry*, vol. 156, no. 7, pp. 1000–1006, 1999.
- [14] L. L. Judd, P. J. Schettler, and H. S. Akiskal, "The prevalence, clinical relevance, and public health significance of sub-threshold depressions," *Psychiatric Clinics of North America*, vol. 25, no. 4, pp. 685–698, 2002.
- [15] J. Angst and R. Sellaro, "Historical perspectives and natural history of bipolar disorder," *Biological Psychiatry*, vol. 48, no. 6, pp. 445–457, 2000.
- [16] M. Olfson, R. C. Kessler, P. A. Berglund, and E. Lin, "Psychiatric disorder onset and first treatment contact in the United States and Ontario," *American Journal of Psychiatry*, vol. 155, no. 10, pp. 1415–1422, 1998.
- [17] M. J. Zaki, J. T. L. Wang, and H. T. T. Toivonen, "Data mining in bioinformatics," *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 2, pp. 198–199, 2003.
- [18] D. Chauhan and V. Jaiswal, "An efficient data mining classification approach for detecting lung cancer disease," in *Proceedings of the 2016 International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, October 2016.
- [19] R. Polikar, "Ensemble learning," *Scholarpedia*, vol. 4, no. 1, p. 2776, 2009.
- [20] D. Greenstein, J. D. Malley, B. Weisinger, L. Clasen, and N. Gogtay, "Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls," *Frontiers in Psychiatry*, vol. 3, 2012.
- [21] Y. T. Jo, S. W. Joo, S. H. Shon, H. Kim, Y. Kim, and J. Lee, "Diagnosing schizophrenia with network analysis and a machine learning method," *International Journal of Methods in Psychiatric Research*, vol. 29, no. 1, 2020.
- [22] H. Yang, J. Liu, J. Sui, G. Pearlson, and V. D. Calhoun, "A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia," *Frontiers in Human Neuroscience*, vol. 4, 2010.
- [23] S. Srinivasagopalan, J. Barry, V. Gurupur, and S. Thankachan, "A deep learning approach for diagnosing schizophrenic patients," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 31, no. 6, pp. 803–816, 2019.
- [24] R. N. Pläschke, E. C. Cieslik, V. I. Müller et al., "On the integrity of functional brain networks in schizophrenia, Parkinson's disease, and advanced age: evidence from connectivity-based single-subject classification," *Human Brain Mapping*, vol. 38, no. 12, pp. 5845–5858, 2017.
- [25] W. H. L. Pinaya, A. Gadelha, O. M. Doyle et al., "Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia," *Scientific Reports*, vol. 6, no. 1, 2016.
- [26] W. H. L. Pinaya, A. Mechelli, and J. R. Sato, "Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: a large-scale multi-sample study," *Human Brain Mapping*, vol. 40, no. 3, pp. 944–954, 2018.
- [27] A. M. Chekroud, R. J. Zotti, Z. Shehzad et al., "Cross-trial prediction of treatment outcome in depression: a machine learning approach," *The Lancet Psychiatry*, vol. 3, no. 3, pp. 243–250, 2016.
- [28] A. Ahmed, R. Sultana, M. T. R. Ullas, M. Begom, M. M. I. Rahi, and M. A. Alam, "A machine learning approach to detect depression and anxiety using supervised learning," in *Proceedings of the 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Gold Coast, Australia, 2019.
- [29] A. Sau and I. Bhakta, "Predicting anxiety and depression in elderly patients using machine learning technology," *Healthcare Technology Letters*, vol. 4, no. 6, pp. 238–243, 2017.
- [30] C. D. Katsis, N. S. Katertsidis, and D. I. Fotiadis, "An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders," *Biomedical Signal Processing and Control*, vol. 6, no. 3, pp. 261–268, 2011.
- [31] A. Sau and I. Bhakta, "Screening of anxiety and depression among seafarers using machine learning technology," *Informatics in Medicine Unlocked*, vol. 16, Article ID 100228, 2019.
- [32] K. Hilbert, U. Lueken, M. Muehlhan, and K. Beesdo-Baum, "Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: a multimodal machine learning study," *Brain and Behavior*, vol. 7, no. 3, Article ID e00633, 2017.
- [33] A. J. Xu, M. A. Flannery, Y. Gao, and Y. Wu, "Machine learning for mental health detection," 2019, <https://digitalcommons.wpi.edu/mqp-all/6732/>.
- [34] V. Rocha-Rego, J. Jogia, A. F. Marquand, J. Mourao-Miranda, A. Simmons, and S. Frangou, "Examination of the predictive value of structural magnetic resonance scans in bipolar disorder: a pattern classification approach," *Psychological Medicine*, vol. 44, no. 3, pp. 519–532, 2013.
- [35] D. Grotegerd, T. Suslow, J. Bauer et al., "Discriminating unipolar and bipolar depression by means of fMRI and pattern classification: a pilot study," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 263, no. 2, pp. 119–131, 2012.
- [36] G. Valenza, M. Nardelli, A. Lanata et al., "Predicting mood changes in bipolar disorder through heartbeat nonlinear dynamics," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1034–1043, 2016.
- [37] J. Mourão-Miranda, J. R. Almeida, S. Hassel et al., "Pattern recognition analyses of brain activation elicited by happy and neutral faces in unipolar and bipolar depression," *Bipolar Disorders*, vol. 14, no. 4, pp. 451–460, 2012.
- [38] G. Roberts, A. Lord, A. Frankland et al., "Functional dysconnection of the inferior frontal gyrus in young people with

- bipolar disorder or at genetic high risk,” *Biological Psychiatry*, vol. 81, no. 8, pp. 718–727, 2017.
- [39] G. Akinci, E. Polat, and O. M. Koçak, “A video based eye detection system for bipolar disorder diagnosis,” in *Proceedings of the 2012 20th Signal Processing and Communications Applications Conference (SIU)*, Mugla, Turkey, April 2012.
- [40] M.-J. Wu, I. C. Passos, I. E. Bauer et al., “Individualized identification of euthymic bipolar disorder using the cambridge neuropsychological test automated battery (CANTAB) and machine learning,” *Journal of Affective Disorders*, vol. 192, pp. 219–225, 2016.
- [41] A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, “Forecasting the onset and course of mental illness with twitter data,” *Scientific Reports*, vol. 7, no. 1, 2017.
- [42] D. Leightley, V. Williamson, J. Darby, and N. T. Fear, “Identifying probable post-traumatic stress disorder: applying supervised machine learning to data from a UK military cohort,” *Journal of Mental Health*, vol. 28, no. 1, pp. 34–41, 2018.
- [43] S. Papini, D. Pisner, J. Shumake et al., “Ensemble machine learning prediction of posttraumatic stress disorder screening status after emergency room hospitalization,” *Journal of Anxiety Disorders*, vol. 60, pp. 35–42, 2018.
- [44] D. Conrad, S. Wilker, A. Pfeiffer et al., “Does trauma event type matter in the assessment of traumatic load?” *European Journal of Psychotraumatology*, vol. 8, no. 1, Article ID 1344079, 2017.
- [45] C. R. Marmar, A. D. Brown, M. Qian et al., “Speech-based markers for posttraumatic stress disorder in US veterans,” *Depression and Anxiety*, vol. 36, no. 7, pp. 607–616, 2019.
- [46] D. Vergyri, B. Knuth, E. Shriberg et al., “Speech-based assessment of PTSD in a military population using diverse feature classes,” in *Proceedings of the INTERSPEECH 2015*, Dresden, Germany, 2015.
- [47] L. E. Salminen, R. A. Morey, B. C. Riedel, N. Jahanshad, E. L. Dennis, and P. M. Thompson, “Adaptive identification of cortical and subcortical imaging markers of early life stress and posttraumatic stress disorder,” *Journal of Neuroimaging*, vol. 29, no. 3, pp. 335–343, 2019.
- [48] D. Rangaprakash, G. Deshpande, T. A. Daniel et al., “Compromised hippocampus-striatum pathway as a potential imaging biomarker of mild-traumatic brain injury and posttraumatic stress disorder,” *Human Brain Mapping*, vol. 38, no. 6, pp. 2843–2864, 2017.
- [49] M. Sumathi and B. Poorna, “Prediction of mental health problems among children using machine learning techniques,” *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 1, 2016.
- [50] A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, “Predicting mental health problems in adolescence using machine learning techniques,” *PLoS One*, vol. 15, no. 4, Article ID e0230389, 2020.
- [51] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, “Machine learning algorithm validation with a limited sample size,” *PLoS One*, vol. 14, no. 11, Article ID e0224365, 2019.
- [52] S. K. Ernala, M. L. Birnbaum, K. A. Candan et al., “Methodological gaps in predicting mental health states from social media,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, UK, May 2019.
- [53] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K. R. Muller, “Evaluating the visualization of what a deep neural network has learned,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, pp. 2660–2673, 2017.
- [54] Y. Dang, Q. Lin, and P. Huang, “AIOps: real-world challenges and research innovations,” in *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, Montreal, Canada, May 2019.
- [55] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [56] K. M. Hermann, T. Kociský, E. Grefenstette et al., “Teaching machines to read and comprehend,” 2015, <http://arxiv.org/abs/1506.03340>.
- [57] S. Graham, C. Depp, E. E. Lee et al., “Artificial intelligence for mental health and mental illnesses: an overview,” *Current Psychiatry Reports*, vol. 21, no. 11, 2019.
- [58] P. Morillo, H. Ortega, D. Chauca, J. Proaño, D. Vallejo-Huanga, and M. Cazares, “Psycho web: a machine learning platform for the diagnosis and classification of mental disorders,” in *Advances in Neuroergonomics and Cognitive Engineering*, pp. 399–410, Springer International Publishing, Berlin, Germany, 2019.
- [59] A. Holzinger, B. Malle, A. Saranti, and B. Pfeifer, “Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI,” *Information Fusion*, vol. 71, pp. 28–37, 2021.
- [60] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, “Machine learning for medical imaging,” *RadioGraphics*, vol. 37, no. 2, pp. 505–515, 2017.
- [61] R. Kemker, A. Abitino, M. McClure, and C. Kanan, “Measuring catastrophic forgetting in neural networks,” 2017, <http://arxiv.org/abs/1708.02072>.
- [62] W. Bosl, A. Tierney, H. Tager-Flusberg, and C. Nelson, “EEG complexity as a biomarker for autism spectrum disorder risk,” *BMC Medicine*, vol. 9, no. 1, 2011.