Article

Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions

Shumaila Aleem [1], **Noor ul Huda** [1], **Rashid Amin** [1,*], **Samina Khalid** [2], **Sultan S. Alshamrani** [3] **and Abdullah Alshehri** [4]

1   Department of Computer Science, University of Engineering and Technology, Taxila 47080, Pakistan; shumaila.aleem@students.uettaxila.edu.pk (S.A.); ulhudan17@gmail.com (N.u.H.)
2   Computer Science and Information Technology Department, Mirpur University of Science and Technology, New Mirpur City 10250, Pakistan; samina.csit@must.edu.pk
3   Department of Information Technology, College of Computer and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia; susamash@tu.edu.sa
4   Department of Information Technology, Al Baha University, Al Bahah 65731, Saudi Arabia; aashehri@bu.edu.sa
*   Correspondence: rashid4nw@gmail.com

Abstract: Over the years, stress, anxiety, and modern-day fast-paced lifestyles have had immense psychological effects on people's minds worldwide. The global technological development in healthcare digitizes the scopious data, enabling the map of the various forms of human biology more accurately than traditional measuring techniques. Machine learning (ML) has been accredited as an efficient approach for analyzing the massive amount of data in the healthcare domain. ML methodologies are being utilized in mental health to predict the probabilities of mental disorders and, therefore, execute potential treatment outcomes. This review paper enlists different machine learning algorithms used to detect and diagnose depression. The ML-based depression detection algorithms are categorized into three classes, classification, deep learning, and ensemble. A general model for depression diagnosis involving data extraction, pre-processing, training ML classifier, detection classification, and performance evaluation is presented. Moreover, it presents an overview to identify the objectives and limitations of different research studies presented in the domain of depression detection. Furthermore, it discussed future research possibilities in the field of depression diagnosis.

Keywords: depression; machine learning (ML); deep learning (DL); regression

1. Introduction

The modern age lifestyle has a psychological impact on people's minds that causes emotional distress and depression [1]. Depression is a prevailing mental disturbance affecting an individual's thinking and mental development. According to WHO, approximately 1 billion people have mental disorders [2] and over 300 million people suffer from depression worldwide [3]. Depression prevails in suicidal thoughts in an individual. Around 800,000 people commit suicide annually. Therefore, it requires a comprehensive response to deal with the burden of mental health issues [4,5]. Depression may harm the socio-economic status of an individual. People suffering from depression are more reluctant to socialize. Counseling and psychological therapies can help fight depression. Machine learning (ML) aims at creating algorithms that are equipped with the ability to train themselves to perceive complex patterns. This ability helps to find solutions to new problems by using previous data and solutions. ML algorithms implement processes with regulated and standardized outcomes [6,7]. Broadly, ML algorithms are categorized into supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning algorithms. The supervised ML algorithms [8] utilize main inputs to predict known values, whereas the unsupervised ML algorithms [9] divulge unidentified patterns and clusters

within the given data. Semi-supervised learning [10] is concerned with the working of systems by combining both labeled and unlabeled data, and it lies between supervised and unsupervised learning. Reinforcement learning [11] is concerned with interpreting the environment to undergo desired actions and exhibiting outcomes through trial and error. The applications of ML techniques in healthcare have proven to be pragmatic as they can process a huge amount of heterogeneous data and provide efficient clinical insights. ML-based approaches provide an efficient understanding of mental conditions and assist mental health specialists in predictive decision making [12]. ML techniques benefit the prediction and diagnosis in the healthcare domain by generating information from unstructured medical data. The prediction outcomes help to identify high-risk medical conditions in patients for early treatments [13]. In mental disorders, ML techniques help arbitrate the potential behavioral biomarkers [14] to assist healthcare specialists in predicting the contingencies of mental disorders and administering effective treatment outcomes. The techniques help the visualization and interpretation of complex healthcare data. The visualization helps develop an effective hypothesis regarding the diagnosis of mental disorders. The traditional clinical diagnostic approach for depression does not accurately identify the depression complexity. The composition of the symptoms related to mental disorders such as depression can easily be detected and anticipated by utilizing ML methods. Therefore, the ML-based diagnostic approach seems to be an efficient choice for predictive analysis. In the healthcare sector, the major domains used for extracting observations associated with mental disorders through ML can be classified as sensors, text, structured data, and multimodal technology interactions [14]. The sensors data can be analyzed using mobile phones and audio signals. The text sources can be extracted through social media platforms, text messages, and clinical records. The structured data constitute the data extracted from standard screening scales, questionnaires, and medical health records. The multimodal technology interactions include data from human interactions with everyday technological equipment, robot, and virtual agents. The ML approaches can be used to assist in diagnosing mental health conditions. The majority of the studies analyze Twitter data [15–17] and sensors data from mobile devices [18,19] for identifying mood disorders. Analyzing textual data can help extract diagnostic information from the individual's psychiatric records [20]. ML approaches can help to predict risk factors in patients with mental disorders. The analysis of sensor data [20], clinical health records [21,22], and text message data [23] can help predict the severity of mental disorders and suicidal behaviors. Various studies have been put forward to aid medical specialists in identifying depression and multiple other mental disorders. The domain of mental disorders comprises a diverse range of mental illnesses. However, this review paper aims attention at the methods presented for the detection of depression. This review paper focuses on elaborating the ML approaches and algorithms used to diagnose and detect depression in individuals. The paper briefly presents the objectives and limitations of the reviewed studies in depression diagnosis, which will help analyze and recognize the best ML approach for a depression diagnosis. The analysis presented in this review paper can help medical specialists and clinicians choose a suitable diagnosis approach for patients with depression. This review paper presents the following: Significant studies extract mental health-related insights. A general model for depression diagnosis involving data extraction, pre-processing, training ML classifier, detection classification, and performance evaluation is considered. An overview of different ML algorithms to diagnose depression by categorizing these depression detection algorithms into three classes, i.e., classification, deep learning, and ensemble. We discussed the limitations of the reviewed studies in the depression diagnosis domain and a better understanding of the choice of the ML approach for depression diagnosis for clinicians and healthcare professionals. Future research possibilities in the domain of depression diagnosis are listed. The organization of the remaining sections of this paper is as follows: Section 2 consists of a brief description of the past studies. The methodology for depression diagnosis is explained in Section 3. Section 4 describes the depression detection model.

Section 5 explains the future direction in the domain of depression diagnosis. Section 6 describes the conclusion of this review.

2. Related Work

Over the years, there have been numerous studies on the use of ML to amplify the scrutiny of mental disorders. In [24], the authors present a history of depression, imaging, and ML approaches. It also provides reviews on researchers that have used imaging and ML to study depression. The algorithms under review are SVM (linear kernel), SVM (nonlinear kernel), and relevance vector regression. Only one mental health domain (MHD) is used to analyze in this survey. This study did not mention depression screening scales, and there is no comprehensive comparison of algorithms. Garcia et al. [25] surveyed mental health monitoring systems (MHMS) using ML and sensor data in mental disorders. This study also analyzed supervised, unsupervised, semi-supervised, transfer, and reinforcement learning which were applied in the domains of mental well-being, including depression, anxiety, bipolar disorder (BD), migraine, and stress. However, the study only presents a brief review of the cases about MHMS and applications. Gao et al. [26] compared ML-based brain imaging classification and prediction research studies for diagnosing. Major depression disorder (MDD) and BD were analyzed, combined with the utilization of the MRI data. SVM, LDA, GPC, DT, RVM, NN, and LR algorithms are under review in this study. However, depression screening scales used in different studies are not mentioned. It only focuses on MDD and BD-based research studies. Gyeongcheol et al. [27] analyzed five ML algorithms; SVM, Gradient Boosting Machine (GBM), RF, Naïve Bayes, and KNN were applied in the domains of mental disorders. It included PTSD, schizophrenia, depression, ASD, and BD studies. This study reviewed the limited number of ML algorithms and did not specify the advantages of using a particular ML approach.

In [28], the authors analyzed Facebook data to detect depression-relevant factors. The Facebook user's data were analyzed using LIWC. Four supervised learning ML approaches were applied to the acquired data: DT, KNN, SVM, and an ensemble model. Experimental results indicated that DT yielded better classification accuracy. Liu et al. [29] presented a brief review of generic AI-based applications for mental disabilities and an illustration of AI-based exploration of biomarkers for psychiatric disorders. The study [30] reviewed three major approaches for brain analysis for psychiatric disorders, magnetic resonance imaging (MRI), electroencephalography (EEG), and kinesics diagnosis, along with five AI methods, Bayesian model, LR, DT, SVM, and DL. In [31], authors have used DL methodology to extract a representation of depression cues in audio and video to detect depression.

This review has introduced the databases and described objective markers for automatic depression estimation (ADE) to sort out and summarize their work. Furthermore, they reviewed the DL methods (DCNN, RNN, and LTMS) for automatic depression detection to extract the representation of depression from audio and video. Finally, they have discussed challenges and promising directions related to the automatic diagnosis of depression using DL approaches. Table 1 illustrate the overview of different studies.

Table 1. Overview of different studies.

Ref.	Year	Area Focused	Mental Health Domain	Algorithms under Review	Limitation
[24]	2016	Studying depression using imaging and ML approaches	MDD	SVM (linear kernel), SVM (nonlinear kernel), relevance vector regression	No comprehensive comparison of algorithms and did not mention depression screening scales.
[25]	2018	Review of research work on mental health monitoring systems (MHMS) based on multimodal sensing and machine learning	Depression, anxiety, bipolar disorder, migraine, and stress	Supervised, unsupervised, semi-supervised, reinforcement, and transfer learning	No extensive review of the defined domains. No comparative evaluation of models or algorithms was presented.
[26]	2018	ML-based classification and prediction studies of MDD combined with MRI data	MDD and BD	SVM, LDA, GPC, DT, RVM, NN, LR	Depression screening scales used in different studies are not mentioned. Only focuses on MDD and BD-based research studies.
[27]	2019	Reviewed different ML techniques and recommends the working of ML methods in the practical world	PTSD, schizophrenia, depression, ASD, bipolar	SVM, GBM, Random Forest, KNN, Naïve Bayes	The limited number of algorithms utilized as compared to other researches.
[28]	2020	Analysis of Facebook data to detect depression-relevant factors using supervised ML algorithms and linguistic approach	MDD	SVM, CNN, DT, KNN, LR, and RF	Limited attributed of LIWC Software. No scope for semi-supervised learning and DL. The study does not identify the individuals but only assesses the Facebook data.
[29]	2020	Review of EEG, MRI, and Kinesics techniques with related AI applications and algorithms	Psychiatric disorders	DL, Naïve Bayes, LR, DT, SVM	No comprehensive comparison of algorithms and only considered classic shallow learning algorithms.
[30]	2021	Extract depression cues from audio and video for automatic depression estimation	MDD, BD, and Other Mood Disorders (OMD)	DCNN, RNN, LTMS	N/A

3. Methodology for Depression Diagnosis

The detection methodology involves a series of processes, including the data extraction, the pre-processing of the extracted data, feature extraction methods for selecting the required set of features for identifying symptoms of depression, and ML classifiers for classifying the input data into defined data categories. This section discusses each of these steps and the different methods and approaches used for implementing each step.

3.1. Pre-Processing Algorithms

(1) Linear Discriminant Analysis (LDA): LDA is a dimensionality reduction approach that removes redundant features by transforming them from a spatial space onto a lower-dimensional space. LDA reduces the dimensions in each dataset, retains the most important features, and achieves higher class separability [31].

(2) Synthetic Minority Oversampling Technique (SMOTE): SMOTE is a statistical oversampling technique to obtain a synthetically class-balanced dataset. It provides a balanced class distribution that develops synthetic patterns from the minority class [32].

(3) Linguistic Inquiry and Word Count (LIWC): LIWC is a text analysis technique for understanding different emotional, subjective, and structural components present in the spoken and written speech patterns [33].

(4) Hidden Markov Model (HMM): HMM is a probabilistic model used to capture and describe information from observable sequential symbols. In HMM, the observed data are modeled as a series of outputs generated by several internal states [34].

3.2. Feature Extraction Methods

Feature selection is a technique in which those features are selected that are the most accurate predictors of the target variable.

(1) SelectKBest: SelectKBest is a feature extraction approach that retains relevant features and drops unwanted features in the given input data. It is a univariate feature selection approach based on the univariate statistical analysis. It combines the univariate statistical test with selecting the K-number of features based on the statistical result between the variables.

(2) Particle Swarm Optimization (PSO): PSO is a computational process for optimizing nonlinear functions by developing the candidate solution in a repetitive pattern based on a defined quality measure. The general concept of the PSO algorithm is inspired by the swarm actions of birds, flocking, and schooling in nature [35].

(3) Maximum Relevance Minimum Redundancy (mRMR): mRMR is a feature selection approach that manages multivariate temporal data without compressing previous data. The algorithm selects features with the most relevant class and the least correlation between redundant classes. It provides significantly improved class predictions in extensive datasets [36].

(4) Boruta: Boruta is a feature selection approach designed around a Random Forest classification. Boruta is used for extracting all the relevant variables by removing less relevant features, using the statistical analysis iteratively [37].

(5) RELIEFF: RELIEFF algorithm is one of the most successful filtering feature selection methods. RELIEFF algorithm is used to eliminate the redundant features [38].

3.3. Supervised Learning Classifiers

In supervised learning, the specific format is used for the training dataset. Each instance is assigned a label. Datasets are labeled as (x, y) belongs to X, Y where x and y denote a data point. The problem is a classification task if the output y belongs to a discrete domain. If the output is a part of the continuous domain, it is a regression task. The tasks predict the value of the dependent attribute from the variables.

3.3.1. Classification

(1) Naïve Bayes Classifier: A Naive Bayes classifier is dependent on applying Bayes' hypothesis with strong independence assumptions. This classifier depends on basic learning strategies assembled by similitudes that utilize Bayes' hypothesis of probability to build ML models, particularly those identified with report order and disease prediction [39].

(2) KNN Classifier: KNN is used for data regression and classification based on the count of k neighbors [40].

(3) Support Vector Machine Classifier (SVM): SVM [41] is a supervised ML model investigating regression analysis and classification data. It also uses the classification for two-group classification problems. SVM are nonparametric classifiers. For the training set, inputs and outputs are paired in SVM. Decision functions are attained through the input–output pairs that classify the input variables into the new and test datasets. i. Multikernel SVM: Multikernel SVM [42] is a feature selection approach based on oversampling and a hybrid algorithm for improving the classification of binary imbalanced classes.

(4) Decision Tree (DT) Classifier: A DT [43] is a tree-like graph used as a decision support tool. It works with discrete-valued parameters and an inductive philosophy for decision tree is "A good decision tree should be as small as possible". Decision Tree Ensembles:

i. Bagging (RF, DF): Bagging is an ensemble algorithm [8]. It adapts various algorithms on different fragments of a training dataset. The predictions from all algorithms are then combined. Random Forest (RF), an extension of bagging, selects the features fragments in random patterns from the given dataset.

ii. Boosting (GBDT, XGBoost): Gradient Boosting is an ensemble classifier used for supervised ML tasks. It considers the individual algorithms and forms a collective model.

3.3.2. Regression

It is used to comprehend the connection between reliant and free factors. It is generally used to make projections, for example, for deal income for a given business. Linear regression and logistical regression are popular regression algorithms.

(1) Logistic Regression: When the dependent variable is dichotomous, logistic regression is the best regression technique to use (binary). Logistic regression is employed to describe and explain the connection between one dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent variables.

(2) Lasso Regression: Lasso regression is a form of shrinkage-based linear regression. Data values are shrunk toward a center; mean in shrinkage. Simple, sparse models are encouraged by the lasso method.

(3) Elastic Net: Elastic net is a regularized linear regression that incorporates two well-known penalties, the L1 and L2 penalty functions.

(4) SVR: Support vector regression (SVR) allows the flexibility to define how much error is acceptable in each model and find an appropriate line to fit the data.

3.3.3. Deep Learning

Deep learning is a type of ML that enhances computers to gain for a fact and comprehend the world as far as a hierarchy of ideas. The hierarchy of ideas permits the computer to learn confounded ideas by building them out of more straightforward ones; a graph of these hierarchies would be many layers deep. In image processing and computer vision with applications, such as scene understanding, clinical image investigation, robotic perception, augmented reality, video surveillance, and image compression, image segmentation is a key idea. Because of the achievement of DL models in a wide scope of vision applications, there has been a generous measure of works pointed toward creating image segmentation approaches utilizing DL models.

Neural Networks:

The neural network is a classifier that stimulates the human brains and neurons; neural networks (NNs) or artificial neural networks (ANN) are based on a collection of process units (e.g., nodes, neurons, or process layers). The processing unit receives signals from other neurons, combines, transforms them, and generates results.

(1) Convolutional Neural Network: ConvNet, also known as CNN, is a deep learning (DL) method that can take an input picture and assign significance (learnable weights and biases) to various aspects/objects in the image, as well as distinguish one from the other. Compared to other classification methods, the amount of pre-processing needed by a ConvNet is much less. While filters are hand-engineered in basic techniques, ConvNets can learn these filters/characteristics with enough training.

(2) Artificial Neural Network (ANN): ANNs depend on the structure of many interaction units. The preparing unit receives signals from different neurons, consolidates, transforms them, and creates an outcome. The cycle units are generally compared to genuine neurons, giving the artificial neural networks.

(3) DNN: An artificial neural network (ANN) having many layers between the input and output layers is known as a deep neural network (DNN) [9]. Various neural networks have different components, but they all have the same components: neurons, synapses, weights, biases, and functions. These components work in the same way as the human brain and can be taught just like any other machine learning algorithm.

(4) DCNN: A deep convolutional neural network (DCNN) comprises many layers of neural networks. Convolutional and pooling layers are usually alternated in most cases. From left to right in the network, the depth of each filter rises. The final level is usually made up of one or more layers that are completely linked.

(5) RNN: Recurrent neural networks (RNNs) are utilized in language modeling applications because input may flow in either direction. For this purpose, long short-term memory is especially useful. Long short-term memory (LSTM) is a recurrent neural network design utilized in deep learning. LSTM contains feedback connections, unlike conventional feedforward neural networks. It can handle large data sequences as well as single data points.

(6) AiME (Novel Model): An artificial intelligence mental evaluation (AiME) framework for detecting symptoms of depression using multimodal deep networks-based human–computer interactive evaluation.

4. Depression Detection Models

Depression is a type of mental illness which brings a serious burden to individuals, families, and society. Conferred by the WHO, depression will be the most common mental illness by 2030 [44]. In difficult situations, depression leads to suicide. Currently, there is no efficient clinical characterization of depression. It makes the diagnosing process restricted and biased. Diagnosing depression is complicated, depending not only on the educational background, cognitive ability, and honesty of the subject to describe the symptoms but also on the experience and motivation of the clinicians. Comprehensive information and thorough clinical training are needed to diagnose the severity of depression accurately [10]. Hence, in recent years, numerous automatic depression estimation (ADE) systems have been introduced to automatically estimate the severity scale of depression by using different ML algorithms. Figure 1 illustrates various ML algorithms for the diagnosis of depression.
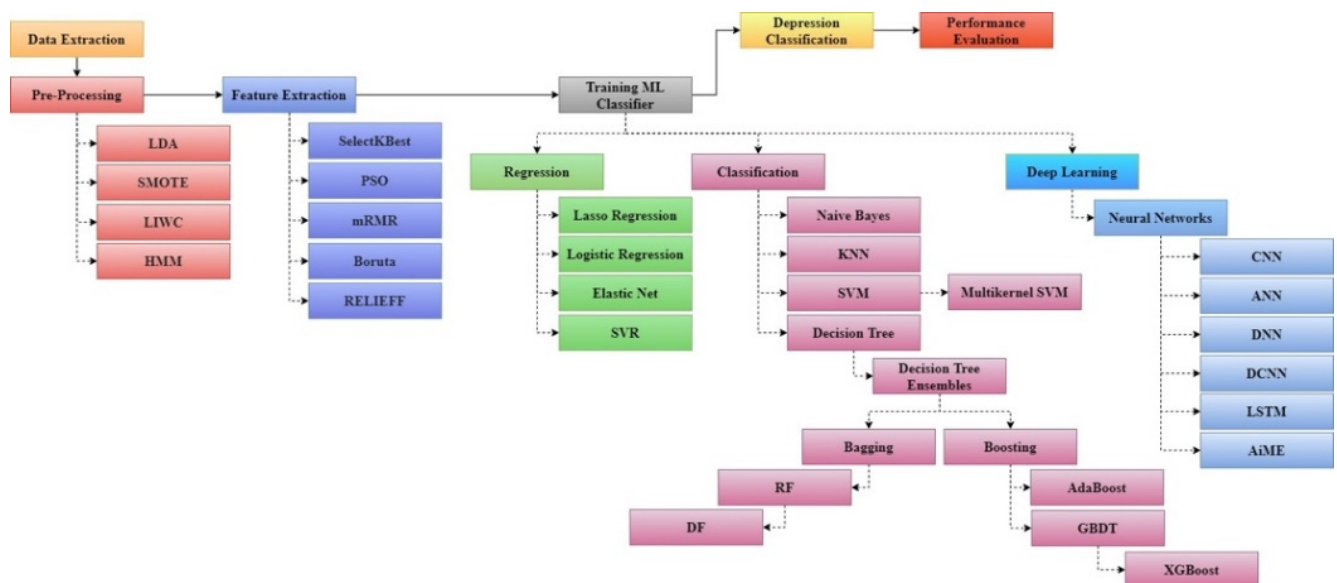

Figure 1. General model for depression diagnosis.

4.1. Classification Models

This section highlights the classification supervised learning models used in several studies for diagnosing depression. A mobile application, Mood Assessment Capable Framework (Moodable), has been presented in [45] to interpret voice samples, data from smartphone and social media handles, and Patient Health Questionnaire (PHQ-9) data for assessment of an individual's mood, mental health, and inferring symptoms of depression by using ML classifiers SVM, KNN, and RF. The framework achieved 76.6% precision

for depression assessment. The authors used six ML classifiers, KNN, Weighted Voting classifier, AdaBoost, Bagging, GB, and XGBoost, in [46], to predict depression. SelectKBest, mRMR, and Boruta feature selection techniques were used for feature extraction. For reducing imbalanced classes, SMOTE was applied. They used a dataset of 604 individuals, including the sociodemographic and psychosocial data and the Burns Depression Checklist (BDC) data, among which 65.73% depression prevalence was identified. The analysis indicated that the AdaBoost classifier achieved the highest classification accuracy of 92.56% when used with the SelectKBest algorithm.

An ML model using the RF algorithm has been implemented for the prognosis of depression among Korean adults in [47]. SMOTE was applied for class balancing between two classes: depression and non-depression. CES-D-11 was used as a depression screening scale where 10-fold cross-validation was utilized to tune the hyperparameters. A total of 6588 Korean citizen's data were included in the study; AUROC value was calculated as 0.870 and achieved an accuracy of 86.20%. However, in this study, biomarkers were not included in the dataset. The authors used three ML algorithms, KNN, RF, and SVM, in [48], to diagnose depression among Bangladeshi students. The study aimed at predicting depression at early stages using related features to avoid drastic incidents. The analysis performed over 577 students' data indicated that the Random Forest algorithm detected the symptoms of depression in the students with 75% accuracy and 60% f-measure.

In [49], ensemble learning and DL approaches have been applied to electroencephalography (EEG) features for detecting depression. Deep Forest (DF) and SVM classifiers were used for feature transformation. Image conversion and CNN were used for feature recognition from the EEG spatial information. The ensemble model with DF and SVM obtained 89.02% classification accuracy and the DL approach achieved 84.75% accuracy. In [50], ML algorithms DT, RF, Naïve Bayes, SVM, and KNN were used to predict stress, anxiety, and depression. The Depression, Anxiety, and Stress Scale questionnaire (DASS 21) analyzed 348 individuals' data. The analysis indicated that Naïve Bayes achieved the highest accuracy of 85.50% for predicting depression. Based on F1 scores, the RF algorithm was more efficient in the case of imbalanced classes. In [51], the author used the sentiment and linguistic analysis with ML to discriminate between depressive and non-depressive social content. RF with RELIEFF feature extractor, LIWC text-analysis tool, and the Hierarchical Hidden Markov Model (HMM) and ANEW scale were used to analyze 4026 social media posts with an accuracy of 90% depressive posts classification, 92% depression degree classification, and 95% depressive communities classification. However, this study takes all depression categories as a single class. Sharma et al. [52] used the XGBoost algorithm on data samples to diagnose mental disorders in the given data. Different sampling techniques were applied to the dataset. The dataset used in this study had imbalanced classes. The study achieved more than 0.90 values for accuracy, precision, recall, and F1 score.

Generalized Anxiety Disorder (GAD) is difficult to perceive and distinguish from major depression (MD) in a clinical framework. In [53], a multi-model ML algorithm was presented to distinguish GAD from MD using structural MRI data and clinical and hormonal information. Conclusively, MRI data provided accumulative data to the GAD classification. However, the sample size and accuracy needed to be increased, and the groups were unbalanced. Xiang et al. [54] used a multikernel SVM with minimum spanning tree (MST) and Kolmogorov–Smirnov test for feature selection. The proposed approach provided a conducive network analysis. A total of 38 MDD patients and 28 healthy controls were included in the dataset. The presented approach achieved 97.54% accuracy. Table 2 presents a comparison of different classification models used for the diagnosis of depression.

Table 2. Comparison of different classification models for depression diagnosis.

Ref.	Objective	Sample Size	Method/ML Classifier	Model Limitation	Depression Screening Scale	Result
[45]	Instantaneous mood assessment using voice samples, mobile and social media data	202 (training), 335 (testing) participant's data	Moodable Application with SVM, KNN, and RF	Not feasible for larger datasets	PHQ-9	76.6% Acc
[46]	Diagnosis of depression using various psychosocial and socio-demographic factors	604 Bangladeshi citizens	KNN, AdaBoost, GB, XGBoost, Bagging, Weighted Voting with SelectKBest, mRMR, Boruta feature selection, and SMOTE	No use of any biological marker and only BDC was considered as ground truth for diagnosis	Burns Depression Checklist (BDC)	92.56% Acc (AdaBoost with SelectKBest)
[47]	An ML-based predictive model for early depression detection	6588 Korean citizens (6067 non-depression and 521 depression)	RF with SMOTE, 10-fold cross-validation, AUROC	Biomarkers were not included in the dataset	CES-D-11	86.20% Acc
[51]	Use of linguistic and sentiment analysis with ML to distinguish depressive and non-depressive social media content	4026 social media posts	RF with RELIEFF feature extractor, LIWC text-analysis tool, and Hierarchical Hidden Markov Model (HMM) and ANEW scale	All depression categories are taken as a single class for classification	Hamilton Depression Rating Scale	Acc% / 90% depressive posts classification / 92% depression degree classification / 95% depressive communities' classification
[52]	XGBoost is less expensive computationally than neural network and easy to implement	11,081	XGBoost	Need to use more datasets that are accepted by other countries and ethnic groups	-	90%
[53]	Classifying GAD and MD subjects based on the incremental value	14 MD, 19 GAD, 24 healthy	SVM	Need to improve accuracy rate, small sample size, and unbalanced classes	STAI-T, PSWQ, BDI, and IUS-12	90.10% Acc
[54]	Improve the accuracy of MDD diagnosis	38 MDD, 28 healthy	Multikernel SVM with MST and Kolmogorov–Smirnov test for feature selection	Limited dataset	-	97.54% Acc
[50]	To identify symptoms of depression, anxiety, and scale using ML algorithms	348 participants	DT, RF, Naïve Bayes, SVM, KNN	Imbalanced classes and smaller dataset	DASS 21	Acc% / 85.50 (NB) / 79.80 (RF) / 77.80 (DT) / 80.30 (SVM) / 72.10 (KNN)
[48]	Predicting depression in university students by identifying related features	577 Bangladeshi undergraduate students	RF, SVM, KNN	Smaller dataset	BDI-II, DASS-21-BV	Acc% / 75 (RF) / 73 (SVM) / 67 (KNN)
[49]	Recognition of depression using transformation of EEG features	EEG data of 14 depression patients, 14 normal subjects	Ensemble and DL model with DF and SVM	Limited dataset	Mini International Neuropsychiatric Interview (MINI)	Acc% / 89.02 (Ensemble model) / 84.75 (DL)

Discussion of Classification Models

The multikernel SVM proposed in [54] with a high-order MST achieved the highest 97.54% MDD classification accuracy among the reviewed studies. The multikernel SVM model provides dynamic changes in the functional association between brain fragments. The integration of multiple kernels can enhance classification. Another model with an efficient classification accuracy was presented in [46], which achieved 92.56% classification accuracy using the AdaBoost with SelectKBest feature selection method and SMOTE for balancing the classes. AdaBoost falls under the category of DT Ensemble. By comparing both the studies [46,54], it can be concluded that in [46], no biomarker was included in the dataset, while in [54], the dataset used was limited and there was no identification of any depression screening scale. Considering the studies [45,48–50,53,54], SVM has been the most used classifier for the detection of depression as it works well on unstructured and high-dimensional data. SVM is also resistant to overfitting. For data with an anonymous and irregular distribution, SVM can be proved to be an efficient algorithm.

Random Forest (RF) is the second most used classifier in the reviewed studies [45,47,48,50,51] as it is a computationally efficient algorithm. In [51], the RF model achieved 90, 95, and 92% accuracy for classifying depressive posts, depressive communities, and depression degrees. RF enhances the classification accuracies of continuous data by reducing the overfitting in decision trees. As RF is based on ensemble learning; it allows determining complex and straightforward functions more accurately. Figure 2 shows the comparison of classification models used for a depression diagnosis.
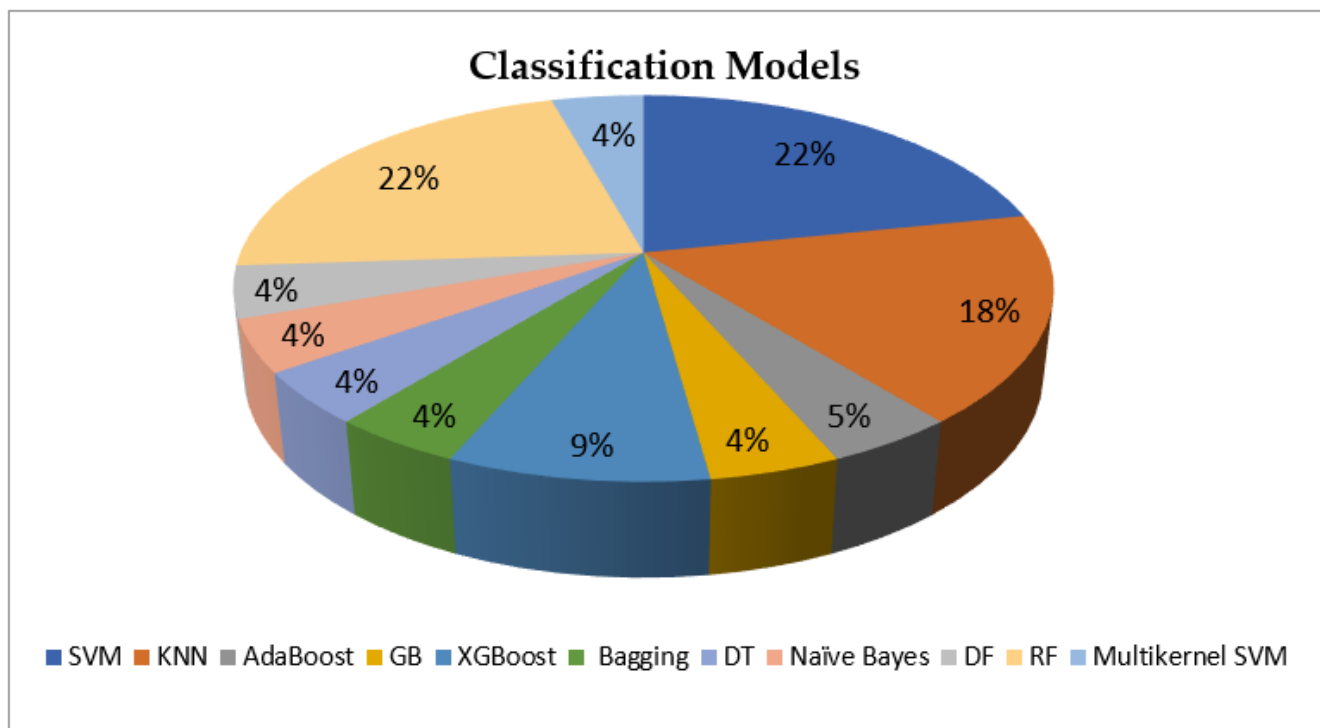

Figure 2. Comparison of classification models for depression diagnosis.

4.2. Deep Learning Models

This section highlights the deep learning models presented in multiple studies to detect depression. An artificial intelligence mental evaluation (AiME) framework [55] has been presented in a study for detecting symptoms of depression using multimodal deep networks-based human–computer interactive evaluation. The framework was applied to audio, video, and speech responses of 671 participants and PHQ-9 data. The authors of [56] discuss the multimodal stress detection using fusion of machine learning algorithms.

In [56], a DL framework based on EEG data have been suggested for the automatic analysis of depression. The framework includes two DL models; one-dimensional convolutional neural network (1DCNN) and a combination of 1DCNN and LSTM model have been utilized. The dataset used in the study contained 30 healthy and 33 MDD patients' EEG data and quantitative information. BDI-II and HADS were used as the assessment scales. The framework achieved an overall classification accuracy of 98.32%. Erguzel, Sayar et al. [57] presented a hybridized methodology using PSO and ANN to distinguish between unipolar and bipolar depression based on EEG recordings. The presented ANN–PSO approach discriminated 31 bipolar and 58 unipolar subjects with 89.89% accuracy. SCID-I, HDRS 17-item version, YMRS, DSM-IV, and HADS were used as the assessment scales. However, this study used limited datasets.

Feng et al. [58] presented the X-A-BiLSTM model for diagnosing depression from social media data. The XGBoost component helped reduce imbalanced classes, and the Attention-BiLSTM neural network component enhanced the classification capacity. The RSDD dataset with approximately 9000 depressed users and 107,000 control users was used in the study. However, no standard screening scale for depression was used in their work. In [59], a novel approach was presented to optimize word embedding for classification. The proposed approach outperformed the previous state-of-the-art models on the RSDD dataset. The comparative evaluation was performed on some DL models for diagnosing depression from tweets on the user level. The experiments were performed on two publicly available datasets, CLPsych 2015 and Bell Let's Talk. Results showed that CNN-based models performed better than RNN-based models. However, the word embedding models did not perform efficiently with larger datasets.

Zogan et al. [59] presented interpretive Multimodal Depression Detection with Hierarchical Attention Network (MDHAN) to detect depressed people on social media. User posts along with Twitter-based multimodal features were considered. The semantic sequence features were captured from the individuals' profiles. MDHAN outperformed other baseline methods. It determined that combining DL with multi-model features can be effective. MDHAN achieved excellent performance and ensured adequate evidence to explain the prediction with an accuracy of 89.5%. However, this study needs to use a standard dataset of Twitter users because the social media data may be vague and can manipulate the experimental outcome. In [60], deep convolutional neural networks (DCNN) are designed to learn deep-learned characteristics from spectrograms and raw voice waveforms in the first place. To improve the depression recognition performance, we suggest using joint fine-tuning layers to merge the raw and spectrogram DCNN.

He and Cao [60] used DCNN to enhance depression classification. DCNN with LLD and MRELBP texture descriptors were applied on 100 training, 100 development, and 100 testing samples. AVEC2013 and AVEC2014 datasets were combined. The results were the MAE of 8.1901 and the RMSE of 9.8874 for the combined dataset. In [61], the authors presented a model for diagnosing mild depression by processing EEG signals using CNN. The model used four functional connectivity metrics (coherence, correlation, PLV, and PLI). The model obtained a classification accuracy of 80.74%. Only functional connectivity matrices are used in the research, and other metrics need to be used for evaluation. Ahmed et al. [62] discussed early depression diagnosis by analyzing posts of Reddit users using a DL-based hybrid model. BiLSTM with Glove, Word2Vec, and Fastext embedding techniques, Meta-Data features, and LIWC were applied on 401 (for testing) and 486 (for training) with 531,453 posts for depression detection. Beck Depression Inventory (BDI) was used as an assessment scale. The proposed model obtained F1 score, precision, and recall of 81, 78, and 86%, respectively. Table 3 presents a comparison of different deep learning models used for the diagnosis of depression.

Table 3. Comparison of deep learning models for depression diagnosis.

Ref.	Objective	Sample Size	Method/ML Classifier	Model Limitation	Depression Screening Scale	Result
[55]	AI-based framework for depression detection with minimum human interaction.	671 US citizens	AiME with multimodal deep networks with LSTM	The behavioral results of participants conducted at a specific time period may be emotionally influenced by an immediate event and not particularly associated with depression.	PHQ-9	Acc: 69.23% / Specificity: 87.77% / Sensitivity: 86.81%
[56]	An EEG-based DL model for diagnosing unipolar depression.	30 healthy controls and 33 MDD patients	1DCNN, 1DCNN with LSTM and 10-fold cross-validation	The process needs a GUI to be used in a clinical environment. Dataset is smaller. The use of anti-depressants, caffeine, and smoking may have negative effects on the classification results of the model.	BDI-II, HADS	1DCNN: 98.32% Acc / 1DCNN with LSTM: 95.97% Acc
[57]	A hybridized a methodology using PSO and ANN to discriminate unipolar and bipolar disorders based on EEG recordings.	89 subjects (31 bipolar and 58 unipolar)	ANN with PSO for feature selection	Smaller dataset.	DSM-IV, SCID-I, HDRS, YMRS	89.89% Acc
[58]	DL based depression detection in imbalanced social media data.	Reddit posts of 9000 users and 107,000 control users	DL model X-A-BiLSTM with XGBoost and Attention-BiLSTM	No use of depression screening scale.	None	Precision: 69% / Recall: 53% / F1: 60%
[63]	Diagnosing depression from Twitter data by using an effective DNN architecture and by optimizing word embedding.	1145 Twitter users	CNN With Max, Multi-Channel CNN, Multi-Channel Pooling CNN, and Bidirectional LSTM with NLTK Tweet tokenizer, Word2Vec word embeddings (Skip-gram, CBOW, Rand)	The word embedding models do not perform efficiently with larger datasets. No use of depression screening scale.	None	Acc% / CLPsych 2015: 87% / Bell Let's Talk: 83%
[59]	Detection of depressed users on social media using hybrid DL model.	4208 users (2159 depressed and 2049 healthy)	Multimodal Depression Detection with Hierarchical Attention Network (MDHAN) with Latent Dirichlet Allocation (LDA) and bidirectional Gated Recurrent Unit (BiGRU) word encoder	No use of a standard dataset of Twitter users; therefore, the social media data used in the research may be vague and can manipulate the experimental outcome.	DSM-IV	89.5% Acc
[60]	Proposed DCNN to boost the depression recognition performance.	100 training, 100 development, 100 testing	Deep convolutional neural networks (DCNN) with LLD and MRELBP texture descriptor	The experimental results are based only on audio data.	BDI-II	MAE: 8.1901 / RMSE: 9.8874

Ref.	Objective	Sample Size	Method/ML Classifier	Model Limitation	Depression Screening Scale	Result
[61]	Diagnosis of mild depression by processing EEG signals using CNN.	24 healthy participants, 24 participants with mild depression	CNN classification model with 24-fold cross-validation and 4 functional connectivity metrics (coherence, correlation, PLV, and PLI)	Only functional connectivity matrices are used in the research, other metrics should be used for evaluation.	BDI-II	80.74% Acc using Coherence functional connectivity matric
[62]	Early depression diagnoses by analyzing posts of Reddit users using a DL-based hybrid model.	401 (for testing) and 486 (for training) with 531,453 posts	BiLSTM with Glove, Word2Vec, and Fastext embedding techniques, Meta-Data features, and LIWC	Imbalanced dataset. The time duration for depression classification is very elongated.	BDI	Word2VecEmbed + Meta feature Set: F1 Score: 0.81 Precision: 0.78 Recall: 0.86

Discussion of Deep Learning Models

The reviewed studies used various DL models with different feature extraction and word embedding techniques in this section. The different DL models presented in [56] showed efficient discrimination between depressed and healthy controls. The 1DCNN achieved the highest classification accuracy of 98.32% and the one-dimensional DCNN with LSTM achieved an accuracy of 95.97%. The DL models automatically discriminate EEG signal patterns.

In the majority of the studies [56,57,61], EEG data have been utilized to diagnose the symptoms of depression in the participants. EEG patterns can help to indicate abnormalities in brain functions and irregular emotional alternations. The EEG signals resemble waves with peaks and valleys with the help of which irregularities can be identified. In [56], a variant of CNN, namely DCNN, was applied over EEG signals to diagnose unipolar depression. In [57], a hybrid model of ANN with PSO algorithm was used to discriminate unipolar and bipolar disorders based on EEG recordings, thereby achieving 89.89% accuracy. In [61], a CNN classification model for diagnosing mild depression by processing the EEG signals was used, and the model achieved 80.74% accuracy using the coherence functional connectivity metric. It can be concluded that EEG-based diagnosis is an efficient and cost-effective method for understanding brain activity and the neural that correlates with social anxiety. Figure 3 presents the comparison of DL models for depression.
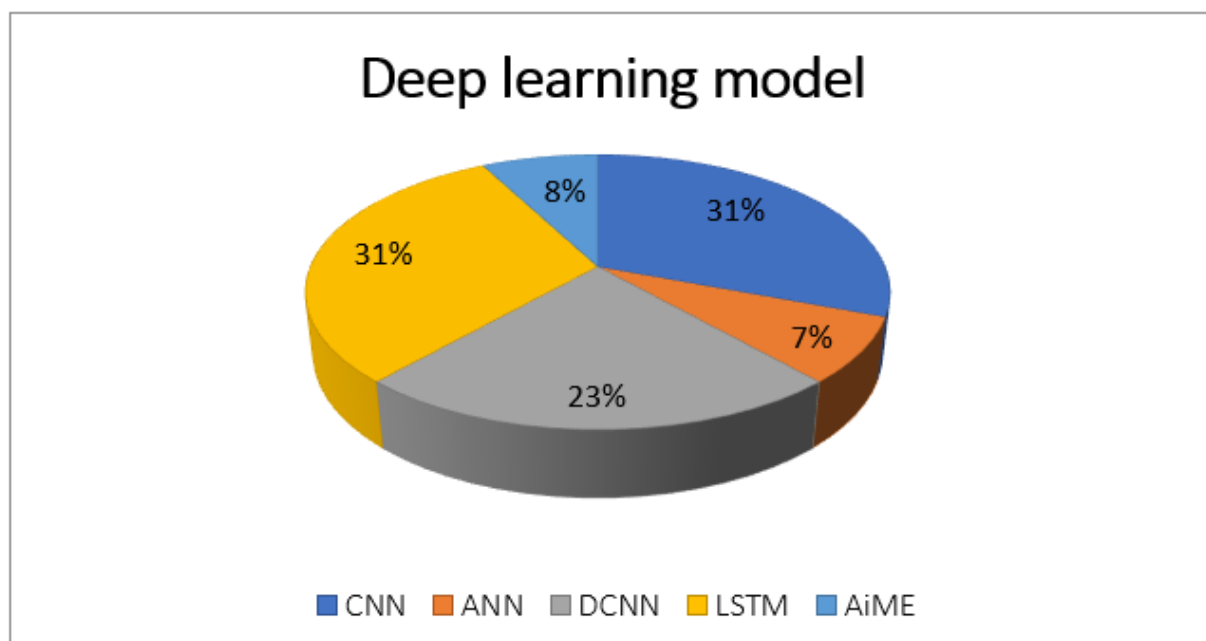

Figure 3. Comparison of deep learning models for depression diagnosis.

4.3. Ensemble Models

This section briefly highlights different ensemble models presented in the reviewed studies for the diagnosis of depression. In [64], ML and statistical models were used to predict clinical depression and MDD among individuals suffering from immune-mediated inflammatory disease (IMID) by identifying patient-reported outcome measures (PROMs). LR, NN, and RF algorithms were used to analyze a dataset of 637 IMID patients. In [65], long short-term memory (LSTM) and six ML models including LR, logistic regression with lasso regularization, RF, gradient boosted decision tree (GBDT), SVM, and deep neural network (DNN) were used. LSTM has been applied to predict the level of different depression risk factors over the course of two years. The dataset contained 1538 data of elderly people in China using the Chinese Longitudinal Healthy Longevity Study (CLHLS).

The results indicated that logistic regression with lasso regularization achieved a higher AUC value than other ML algorithms.

Tao, Chi et al. [66] proposed an ensemble binary classifier to analyze health survey data against ground truth from the SF-20 Quality of Life scales. With ensemble model (DT, AAN, KNN, SVM) applied on the NHANES dataset, the classifier demonstrated an F1 score of 0.976 in the prediction, without any incorrectly identified depression instances. This study has some limitations; the need to use rich online social media sources for feature extraction and dataset range is not defined. Karoly and Ruehlman [67] proposed an algorithm to distinguish between MDD and BD patients based on clinical variables. LR with Elastic Net and XGBoost were applied on 103 MDD and 52 BD patients and achieved an accuracy of 78% for LR with Elastic Net model. There are some limitations in this paper such as the small and unbalanced sample, lack of external sample validation, some misclassifications of classes, and a limited range of evaluation features.

Zhao, Feng et al. [68] evaluated the depression status of Chinese recruits using ML algorithms. NN, SVM, and DT were applied on 1000 participants and achieved 86, 86, and 73% accuracy for NN, SVM, and DT. BD-II was used as an assessment scale. This study needs to include complex socio-demographic and career variables into the model. Ji et al. [69] diagnosed bipolar disorder among Chinese by developing a BDCC using ML algorithms. SVR, RF, LASSO, LR, and LDA were applied on 255 MDD, 360 BPD, and 228 healthy sample data. The experiments obtained an accuracy of 92% for MDD and 92% for BPD detection. However, this model requires large datasets and needs to enhance its cross-sectional nature. Table 4 presents a comparison of different ensemble models used for the diagnosis of depression.

Table 4. Comparison of different ensemble models for depression diagnosis.

Ref.	Objective	Sample Size	Method/ML Classifier	Model Limitation	Depression Screening Scale	Result
[64]	To utilize ML to predict MDD and anxiety disorder in IMID patients.	637 IMID patients	LR, NN, RF with AUROC with 10-fold cross-validation, Brier scores	Participants in the study have different IMID conditions. No use of PROM instruments and separate testing dataset.	SCID, PROMs	For LR: AUC: 0.90 Brier score: 0.07 For NN: AUC: 0.90 Brier score: 0.07 For RF: AUC: 0.91 Brier score: 0.07
[65]	To predict depression among elderly people of China using ML.	1538 elderly Chinese participants	LR, LR with lasso regularization, RF, GBDT, SVM, and DNN with LSTM	Retrospective waves in the LSTM need to be increased. No use of depression screening scale.	N/A	0.629 AUC (LR with lasso regularization)
[66]	MDD diagnosis using an ensemble binary classifier.	NHANES dataset	Ensemble model with DT, AAN, KNN, SVM	No use of rich online social media sources for feature extraction. Dataset range is not defined.	PHQ-9, SF-20 QOLS	95.4% Acc
[67]	Development of an algorithm to distinguish between MDD and bipolar disorder (BD) patients based on clinical variables.	103 MDD and 52 BD patients	LR with Elastic Net and XGBoost	Small and unbalanced sample. Lack of external sample validation. Some misclassifications of classes. Lesser evaluation features.	Brazilian version of TCI, BDI, STAI, PANAS	78% Acc for LR with Elastic Net Model

Table 4. *Cont.*

Ref.	Objective	Sample Size	Method/ML Classifier	Model Limitation	Depression Screening Scale	Result
[68]	Evaluating the depression status of Chinese recruits using ML.	1000 participants	NN, SVM, DT	Need to include complex socio-demographic variables and career variables into the model.	BDI-II	Acc%
						86 (SVM)
						86 (NN)
						73 (DT)
[69]	Diagnosis of bipolar disorder among Chinese by developing a Bipolar Diagnosis Checklist in Chinese (BDCC) by using ML algorithms.	255 MDD, 360 BPD, 228 healthy	SVR, RF, LASSO, LR, and LDA	Require large datasets and need to enhance its cross-sectional nature.	N/A	92% (MDD) 92% (BPD)

Discussion of Ensemble Models

Among the reviewed studies, ensemble models [66] obtained the highest accuracy of 95.4%. In this study, the NHANES dataset is used for evaluation; the predicted model just predicts the 4% cases wrongly. The ensemble model achieved F1 measure, accuracy, and precision of 97, 95, and 95%, respectively, on the whole dataset. It also shows that the ensemble method for identifying depression on a partial dataset is stable and resilient. The method and experiment showed that combining a classification methodology with binary ground truth may provide better prediction results than baseline standards. The ensemble technique is a straightforward approach similar to the bagging and major voting ensemble methods. Using five machine learning algorithms and Chinese multicenter cohort data, the ensemble model described in [69] obtained the second-highest classification accuracy of 92 percent. The higher AUC obtained in this study, compared to other studies, shows the research's acceptance and the validity of the Chinese version of the BDCC. In addition, the BDCC cuts the time it takes to gather clinical data in half. The ADE takes more than 30 min to complete, while the BDCC takes 10–15 min. The present findings show that the BDCC is just as reliable as the previous form, but it is much easier to deploy. Considering the studies [64,65,67,69], regression has been the most used ML technique for the detection of depression. Regression is simple to implement and easier to interpret the output coefficients. Regression is susceptible to overfitting, but it can be avoided using dimensionality reduction techniques, regularization (L1 and L2) techniques, and cross-validation.

5. Future Research Possibilities

We propose some possible future study directions in this part, based on the review of prior research in the preceding section.

(1) A larger data sample is required:

The majority of prior depression detection research utilized a small sample size. A small sample size is useful for building a prediction model, while a bigger sample size is important for constructing a more accurate model that works well throughout the population. When a large sample size is used to train a model, it allows for a greater diversity of depressed patients to be included, perhaps leading to models with real therapeutic value. When a few studies use bigger datasets, the methods will most likely alter and show more developed approval metrics. The k-fold cross-validation technique, in particular, may be employed with higher k-values to allow for larger test sets on which to test prediction models and increase generalizability.

(2) Learning method(s):

Various learning techniques give a better outcome in different situations; therefore, choosing the right one is crucial. Unlabeled data may sometimes help develop a prediction

model for a large sample size with little data. As a result, the first step is to determine if the incoming data are labeled, unlabeled, or a combination of labeled and unlabeled data. As a result, employing an unsupervised, supervised, or semi-supervised learning technique will be determined. The second phase is dependent on the learning method's objective, which must be addressed. The last stage is to identify whether the input is linear or nonlinear; linear data are helpful when the dataset is small to prevent overfitting, whereas nonlinear data are important when the dataset is big. The last step is to choose a learning technique to limit the options. The technique for picking the best learning method is to assess various factors such as complexity, flexibility, computation time, optimization ability, and so on, and then choose the best one. If you have too many learning method choices, evaluate the performance of each technique on the provided data; if you just have a few, simply change the default model to make it more appropriate for learning the given data.

(3)   Clinical application:

Long-term, creating a predictive model aims to find a method that can improve accuracy. However, such a scenario is unlikely to arise in the next few years, since SVM and a few other supervised learning algorithms are presently trustworthy and seem to be around in this area of research. Regardless, after a sufficiently strong method has been thoroughly authorized via preliminary considerations, showing its efficacy, and determining whether it will benefit patients or not, its progression to clinical preliminaries will be critical. Future clinical trials should ensure that machine learning methods efficiently identify depressed individuals who are unlikely to respond to the current specialist under investigation. Clinicians' use of this information improves patient outcomes (for example, diminished inactivity among determination and reduction).

(4)   Collaboration of research groups:

With the significant progress among different disciplines, collaboration with other disciplines is crucial for ADE. For affective computing, relevant fields include psychology, physiology, computer science, ML, etc. Thus, researchers should borrow each other's strengths to promote ADE's advances. For audio-based ADE, the deep models only represent the depression scale from audios. The deep models capture patterns only from facial expressions specific to video-based ADE. Notably, physiological signals also contain significant information closely related to depression estimation. Accordingly, different researchers should study together to build multimodal-based DL approaches for clinical application.

(5)   Availability of databases:

Because of the sensitivity of depression data, it is difficult to gain various data for estimating the scale of depression. Hence, the availability of data is a major issue. First, as opposed to the facial expression recognition task, database availability is scarce up to the present day. Given the literature review, one can note that the widely used depression databases are AVEC2013, AVEC2014, and DAIC-WOZ. Notably, AVEC2014 is a subset of AVEC2013. Second, there is no multimodal (i.e., audio, video, text, physiological signals) database to learn comprehensive depression representations for ADE. The existing databases consist of two or three modalities. Though the DAIC database comprises three modalities (audiovisual and text), the organizer has not provided the original videos of DAIC, leading to a certain inconvenience for ADE. Third, the limited size of the datasets limits the research in depression prediction, especially when using DL technologies. For instance, AVEC2013 only contains 50 samples for training, development, and test set. Effective methods to augment the limited amount of annotated data are called to address this bottleneck. Fourth, the criteria for data collection should be standardized. At present, different organizers adopt a range of conditions, equipment, and configurations to collect multimodal data.

6. Conclusions

The ML approaches can be used to assist in diagnosing mental health conditions. PTSD, schizophrenia, depression, ASD, and bipolar diseases lie in the domains of mental disorders. Social media data, clinical health records, and mobile devices sensors data can be analyzed to identify mood disorders. In this paper, we surveyed state-of-the-art research studies on the diagnosis of depression using ML-based approaches. The purpose of this review paper is to provide information about basic concepts of ML algorithms frequently used in the mental health domain, specifically for depression and their practical application. Among the reviewed studies, SVM has been the most used classifier for detecting depression as it works well with unstructured and high-dimensional data. SVM is also resistant to overfitting. SVM can be proved to be an efficient algorithm for data with an anonymous and irregular distribution. As anticipated, most of the SVM classifiers developed in the articles had a high accuracy of greater than 75%. Because data in the mental health area are scarce, SVM outperforms other machine learning methods for diagnosis. We discussed some of the MHMS's research difficulties and potential advancements in mental health and depression. According to the research reviewed, applications based on machine learning provide a significant potential for progress in mental healthcare, including the prediction of outcomes and therapies for mental illnesses and depression.

References

1. Vitriol, V.; Cancino, A.; Weil, K.; Salgado, C.; Asenjo, M.A.; Potthoff, S. Depression and psychological trauma: An overview integrating current research and specific evidence of studies in the treatment of depression in public mental health services in chile. *Depress. Res. Treat.* **2014**, *2014*, 608671. [CrossRef] [PubMed]
2. World Mental Health Day: An Opportunity to Kick-Start a Massive Scale-Up in Investment in Mental Health. Available online: https://www.who.int/news/item/27-08-2020-world-mental-health-day-an-opportunity-to-kick-start-a-massive-scale-up-in-investment-in-mental-health (accessed on 20 February 2022).
3. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2018**, *392*, 1789–1858. [CrossRef]
4. Strunk, D.R.; Pfeifer, B.J.; Ezawa, I.D. Depression. In *Handbook of Cognitive Behavioral Therapy: Applications*; Wenzel, A., Ed.; American Psychological Association: Washington, DC, USA, 2021; pp. 3–31. [CrossRef]
5. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. Available online: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf (accessed on 20 February 2022).
6. ÇELİK, Ö.; Altunaydin, S.S. A research on machine learning methods and its applications. *J. Educ. Technol. Online Learn.* **2018**, *1*, 25–40. [CrossRef]
7. Shalev-Shwartz, S.; Ben-David, S. *Decision Trees. Understanding Machine Learning*; Cambridge University Press: Cambridge, UK, 2014.
8. Kotsiantis, S.B. Supervised machine learning: A review of classification techniques. *Informatica* **2007**, *31*, 249–268.
9. Alloghani, M.A.; Al-Jumeily, D.; Mustafina, J.; Hussain, A.; Aljaaf, A.J. A systematic review on supervised and unsupervised machine learning algorithms for data science. In *Supervised and Unsupervised Learning for Data Science*; Berry, M., Mohamed, A., Yap, B., Eds.; Springer: Cham, Switzerland, 2020.
10. Van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440. [CrossRef]
11. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT press: Cambridge, MA, USA, 2018.
12. Dwyer, D.B.; Falkai, P.; Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* **2018**, *14*, 91–118. [CrossRef]

13. Sidey-Gibbons, J.A.; Sidey-Gibbons, C.J. Machine learning in medicine: A practical introduction. *BMC Med. Res. Methodol.* **2019**, *19*, 1–18. [CrossRef]
14. Thieme, A.; Belgrave, D.; Doherty, G. Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Trans. Comput.-Hum. Interact.* **2020**, *27*, 34. [CrossRef]
15. Chen, X.; Sykora, M.D.; Jackson, T.W.; Elayan, S. What about mood swings: Identifying depression on twitter with temporal measures of emotions. In Proceedings of the the Web Conference, Lyon, France, 23–27 April 2018; pp. 1653–1660.
16. Joshi, D.J.; Makhija, M.; Nabar, Y.; Nehete, N.; Patwardhan, M.S. Mental health analysis using deep learning for feature extraction. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, Goa, India, 11–13 January 2018; pp. 356–359.
17. Aldabbas, H.; Albashish, D.; Khatatneh, K.; Amin, R. An Architecture of IoT-Aware Healthcare Smart System by Leveraging Machine Learning. *Int. Arab J. Inf. Technol.* **2022**, *19*, 160–172. [CrossRef]
18. Amin, R.; Al Ghamdi, M.A.; Almotiri, S.H.; Alruily, M. Healthcare Techniques Through Deep Learning: Issues, Challenges and Opportunities. *IEEE Access* **2021**, *9*, 98523–98541.
19. Morshed, M.B.; Saha, K.; Li, R.; D'Mello, S.K.; De Choudhury, M.; Abowd, G.D.; Plötz, T. Prediction of mood instability with passive sensing. *ACM* **2019**, *3*, 1–21. [CrossRef]
20. Diederich, J.; Al-Ajmi, A.; Yellowlees, P. Ex-ray: Data mining and mental health. *Appl. Soft Comput.* **2007**, *7*, 923–928. [CrossRef]
21. Adamou, M.; Antoniou, G.; Greasidou, E.; Lagani, V.; Charonyktakis, P.; Tsamardinos, I. Mining free-text medical notes for suicide risk assessment. In Proceedings of the 10th Hellenic Conference on Artificial Intelligence, Patras, Greece, 9–12 July 2018; pp. 1–8.
22. Tran, T.; Phung, D.; Luo, W.; Harvey, R.; Berk, M.; Venkatesh, S. An integrated framework for suicide risk prediction. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 1410–1418.
23. Nobles, A.L.; Glenn, J.J.; Kowsari, K.; Teachman, B.A.; Barnes, L.E. Identification of imminent suicide risk among young adults using text messages. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–11.
24. Patel, M.J.; Khalaf, A.; Aizenstein, H.J. Studying depression using imaging and machine learning methods. *NeuroImage Clin.* **2016**, *10*, 115–123. [CrossRef] [PubMed]
25. Garcia-Ceja, E.; Riegler, M.; Nordgreen, T.; Jakobsen, P.; Oedegaard, K.J.; Tørresen, J. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive Mob. Comput.* **2018**, *51*, 1–26. [CrossRef]
26. Gao, S.; Calhoun, V.D.; Sui, J. Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neurosci. Ther.* **2018**, *24*, 1037–1052. [CrossRef] [PubMed]
27. Cho, G.; Yim, J.; Choi, Y.; Ko, J.; Lee, S.-H. Review of Machine Learning Algorithms for Diagnosing Mental Illness. *Psychiatry Investig.* **2019**, *16*, 262–269. [CrossRef] [PubMed]
28. Mahdy, N.; Magdi, D.A.; Dahroug, A.; Rizka, M.A. Comparative Study: Different Techniques to Detect Depression Using Social Media. In *Internet of Things-Applications and Future*; Springer: Singapore, 2020; pp. 441–452. [CrossRef]
29. Liu, G.-D.; Li, Y.-C.; Zhang, W.; Zhang, L. A Brief Review of Artificial Intelligence Applications and Algorithms for Psychiatric Disorders. *Engineering* **2020**, *6*, 462–467. [CrossRef]
30. He, L.; Niu, M.; Tiwari, P.; Marttinen, P.; Su, R.; Jiang, J.; Guo, C.; Wang, H.; Ding, S.; Wang, Z.; et al. Deep learning for depression recognition with audiovisual cues: A review. *Inf. Fusion* **2021**, *80*, 56–86. [CrossRef]
31. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A.E. Linear discriminant analysis: A detailed tutorial. *AI Commun.* **2017**, *30*, 169–190. [CrossRef]
32. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
33. Wegner, D.M. *The Illusion of Conscious Will*; MIT Press: Cambridge, MA, USA, 2017.
34. Rabiner, L.; Juang, B.-H. *Fundamentals of Speech Recognition*; Prentice Hall: Hoboken, NJ, USA, 1993.
35. Koohi, I.; Groza, V.Z. Optimizing particle swarm optimization algorithm. In Proceedings of the 2014 IEEE 27th Canadian conference on Electrical and Computer Engineering (CCECE), Toronto, ON, Canada, 4–7 May 2014.
36. Bugata, P.; Drotar, P. On some aspects of minimum redundancy maximum relevance feature selection. *Sci. China Inf. Sci.* **2020**, *63*, 1–15. [CrossRef]
37. Kursa, M.B.; Jankowski, A.; Rudnicki, W.R. Boruta—A system for feature selection. *Fundam. Inform.* **2010**, *101*, 271–285. [CrossRef]
38. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [CrossRef] [PubMed]
39. Rish, I. An empirical study of the naive Bayes classifier. *IJCAI* **2001**, *3*, 41–46.
40. Cunningham, P.; Delany, S.J. k-Nearest neighbour classifiers: 2nd Edition (with Python examples). *arXiv* **2020**, arXiv:2004.04523.
41. Cristianini, N.; Ricci, E. *Encyclopedia of Algorithms*; Springer: Boston, MA, USA, 2008.
42. Chen, Z.; Li, J. A multiple kernel support vector machine scheme for simultaneous feature selection and rule-based classification. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Nanjing, China, 22–25 May 2007; pp. 441–448.
43. Patel, H.H.; Prajapati, P. Study and Analysis of Decision Tree Based Classification Algorithms. *Int. J. Comput. Sci. Eng.* **2018**, *6*, 74–78. [CrossRef]

44. Mathers, C.D.; Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* **2006**, *3*, e442. [CrossRef]

45. Miyajima, A.; Tanaka, M.; Itoh, T. Stem/progenitor cells in liver development, homeostasis, regeneration, and reprogramming. *Cell Stem Cell* **2014**, *14*, 561–574. [CrossRef]

46. Zulfiker, M.S.; Kabir, N.; Biswas, A.A.; Nazneen, T.; Uddin, M.S. An in-depth analysis of machine learning approaches to predict depression. *Curr. Res. Behav. Sci.* **2021**, *2*, 100044. [CrossRef]

47. Na, K.-S.; Cho, S.-E.; Geem, Z.W.; Kim, Y.-K. Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm. *Neurosci. Lett.* **2020**, *721*, 134804. [CrossRef]

48. Choudhury, A.A.; Khan, R.H.; Nahim, N.Z.; Tulon, S.R.; Islam, S.; Chakrabarty, A. Predicting Depression in Bangladeshi Undergraduates using Machine Learning. In Proceedings of the 2019 IEEE Region 10 Symposium (TENSYMP), Kolkata, India, 7–9 June 2019; pp. 789–794. [CrossRef]

49. Li, X.; Zhang, X.; Zhu, J.; Mao, W.; Sun, S.; Wang, Z.; Xia, C.; Hu, B. Depression recognition using machine learning methods with different feature generation strategies. *Artif. Intell. Med.* **2019**, *99*, 101696. [CrossRef] [PubMed]

50. Priya, A.; Garg, S.; Tigga, N.P. Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. *Procedia Comput. Sci.* **2020**, *167*, 1258–1267. [CrossRef]

51. Fatima, I.; Mukhtar, H.; Ahmad, H.F.; Rajpoot, K. Analysis of user-generated content from online social communities to characterise and predict depression degree. *J. Inf. Sci.* **2018**, *44*, 683–695. [CrossRef]

52. Sharma, A.; Verbeke, W.J.M.I. Improving Diagnosis of Depression with XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081). *Front. Big Data* **2020**, *3*, 15. [CrossRef]

53. Hilbert, K.; Lueken, U.; Muehlhan, M.; Beesdo-Baum, K. Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: A multimodal machine learning study. *Brain Behav.* **2017**, *7*, e00633. [CrossRef]

54. Guo, H.; Qin, M.; Chen, J.; Xu, Y.; Xiang, J. Machine-Learning Classifier for Patients with Major Depressive Disorder: Multifeature Approach Based on a High-Order Minimum Spanning Tree Functional Brain Network. *Comput. Math. Methods Med.* **2017**, *2017*, 4820935. [CrossRef]

55. Nagpal, G.; Chaudhary, K.; Agrawal, P.; Raghava, G.P.S. Computer-aided prediction of antigen presenting cell modulators for designing peptide-based vaccine adjuvants. *J. Transl. Med.* **2018**, *16*, 181. [CrossRef]

56. Mumtaz, W.; Qayyum, A. A deep learning framework for automatic diagnosis of unipolar depression. *Int. J. Med Inform.* **2019**, *132*, 103983. [CrossRef]

57. Erguzel, T.T.; Sayar, G.H.; Tarhan, N. Artificial intelligence approach to classify unipolar and bipolar depressive disorders. *Neural Comput. Appl.* **2016**, *27*, 1607–1616. [CrossRef]

58. Cong, Q.; Feng, Z.; Li, F.; Xiang, Y.; Rao, G.; Tao, C. XA-BiLSTM: A deep learning approach for depression detection in imbalanced data. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018; pp. 1624–1627.

59. Zogan, H.; Wang, X.; Jameel, S.; Xu, G. Depression detection with multi-modalities using a hybrid deep learning model on social media. *CoRR* **2020**, abs/2007.02847.

60. He, L.; Cao, C. Automated depression analysis using convolutional neural networks from speech. *J. Biomed. Inform.* **2018**, *83*, 103–111. [CrossRef] [PubMed]

61. Li, X.; La, R.; Wang, Y.; Hu, B.; Zhang, X. A Deep Learning Approach for Mild Depression Recognition Based on Functional Connectivity Using Electroencephalography. *Front. Neurosci.* **2020**, *14*, 192. [CrossRef] [PubMed]

62. Shah, F.M.; Ahmed, F.; Joy, S.K.S.; Ahmed, S.; Sadek, S.; Shil, R.; Kabir, H. Early Depression Detection from Social Network Using Deep Learning Techniques. In Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 5–7 June 2020; pp. 823–826. [CrossRef]

63. Orabi, A.H.; Buddhitha, P.; Orabi, M.H.; Inkpen, D. Deep learning for depression detection of twitter users. In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, New Orleans, LA, USA, 5 June 2018; pp. 88–97.

64. Tennenhouse, L.G.; Marrie, R.A.; Bernstein, C.N.; Lix, L.M. Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease. *J. Psychosom. Res.* **2020**, *134*, 110126. [CrossRef]

65. Su, D.; Zhang, X.; He, K.; Chen, Y. Use of machine learning approach to predict depression in the elderly in China: A longitudinal study. *J. Affect. Disord.* **2021**, *282*, 289–298. [CrossRef] [PubMed]

66. Tao, X.; Chi, O.; Delaney, P.J.; Li, L.; Huang, J. Detecting depression using an ensemble classifier based on Quality of Life scales. *Brain Inform.* **2021**, *8*, 2. [CrossRef]

67. Karoly, P.; Ruehlman, L.S. Psychological "resilience" and its correlates in chronic pain: Findings from a national community sample. *Pain* **2006**, *123*, 90–97. [CrossRef] [PubMed]

68. Zhao, M.; Feng, Z. Machine Learning Methods to Evaluate the Depression Status of Chinese Recruits: A Diagnostic Study. *Neuropsychiatr. Dis. Treat.* **2020**, *16*, 2743–2752. [CrossRef]

69. Ma, Y.; Ji, J.; Huang, Y.; Gao, H.; Li, Z.; Dong, W.; Zhou, S.; Zhu, Y.; Dang, W.; Zhou, T.; et al. Implementing machine learning in bipolar diagnosis in China. *Transl. Psychiatry* **2019**, *9*, 305. [CrossRef]