Machine Learning-Based Prediction of Depressive Disorders via Various Data Modalities: A Survey

***Abstract*—Depression, a pervasive mental health disorder, has substantial impacts on both individuals and society. The conventional approach to predicting depression necessitates substantial collaboration between health care professionals and patients, leaving room for the influence of subjective factors. Consequently, it is imperative to develop a more efficient and accessible prediction methodology for depression. In recent years, numerous investigations have delved into depression prediction techniques, employing diverse data modalities and yielding notable advancements. Given the rapid progression of this domain, the present article comprehensively reviews the major breakthroughs in depression prediction, encompassing multiple data modalities such as electrophysiological signals, brain imaging, audiovisual data, and text. By integrating depression prediction methods from various data modalities, it offers a comparative assessment of their advantages and limitations, providing a well-rounded perspective on how different modalities can complement each other for more accurate and holistic depression prediction. The survey begins by examining commonly used datasets, evaluation metrics, and methodological frameworks. For each data modality, it systematically analyzes traditional machine learning methods alongside the increasingly prevalent deep learning approaches, providing a comparative assessment of detection frameworks, feature representations, context modeling, and training strategies. Finally, the survey culminates with the identification of prospective avenues that warrant further exploration.**

***Index Terms*—Depression Prediction, Machine Learning, Electrophysiological Signals, Brain Imaging, Audiovisual Data, Text.**

I. INTRODUCTION

DEPRESSIVE disorders, also known as depression, is a prevalent mental disorder [1]–[3], affecting over 280 million individuals globally [4]. Characterized by persistent sadness, diminished interest, and impaired daily functioning, it can severely impact patients' quality of life [5]–[7]. In severe cases, depression may trigger self-harm or suicidal ideation [8]–[10], with an estimated twelve-fold increase in suicide risk [11]. Undoubtedly, the consequences of depression are profound.

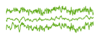
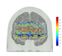
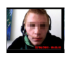
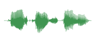
To alleviate the damage caused by depression, an effective prediction is essential. However, four primary challenges persist in clinical depression prediction: (1) *Inadequate objective measures:* Unlike many physiological disorders, depression lacks precise biomarkers, leading clinicians to rely largely on psychological questionnaires and patient self-reports [12]. This reliance can introduce bias, as individuals may provide inaccurate information due to personal biases, social expectations, or challenges in recalling past experiences [13]–[15]. (2) *Scarce medical resources:* The diagnosis and treatment of depression face significant shortages of medical professionals and resources. Approximately 75% of patients do not receive timely diagnosis and treatment, particularly in low- and middle-income countries [16]. Moreover, inefficient and lengthy face-to-face consultations with psychiatrists exacerbate psychiatric care resource constraints [17], [18]. (3) *Financial barriers:* For many depressive individuals, financial limitations restrict access to consistent and adequate care [19]. Mental health services are often expensive, even for those with insurance or financial assistance. (4) *Social stigma:* Public misconceptions frequently lead to resistance or avoidance of mental health care. Individuals may fear altered treatment by family and friends upon disclosing mental health concerns or worry about potential negative repercussions at work [20], [21].

To address the challenges in depression prediction, researchers have recently proposed machine learning-based approaches [22]–[31], which have further strengthened the theoretical foundation in this field. These techniques enable the automatic extraction of relevant features from depression data, unveiling the intrinsic links between clinical symptoms and physiological signals, thereby becoming a crucial direction for depression prediction [32]. The following advantages arise from machine learning-based depression prediction: (1) *High efficiency:* Utilizing machine learning technology to analyze depressed patients' physiological signals and generate objective predictive indicators allows for expedited and effective prediction. (2) *Auxiliary Assistance:* Automatic depression prediction through machine learning methods can effectively complement traditional depression scale screenings, thus reducing the workload of medical professionals. However, despite their robust performance in various tasks, machine learning models are susceptible to biases. This issue arises because these models are trained on historical data, which, if biased, can propagate similar biases in model predictions. Therefore, it is crucial to carefully evaluate and validate machine learning models in real-world settings.

Early studies on depression prediction primarily relied on traditional machine learning methods, achieving some degree of success with small-scale datasets [33]–[36]. With the rise of deep learning, large-scale data and more powerful computational resources have driven significant progress in this field [37]–[43]. Several review papers to date have summarized recent advances in depression prediction, with some, such as [44]–[53], focusing on specific modalities or methodologies. While [54] encompasses a broad range of modalities and machine learning methods, yet its analysis is limited to basic statistical summaries and lacks systematic review and in-depth exploration. Additionally, it only includes studies up to mid-2021, excluding significant recent advancements, as shown in Fig. 1. Thus, a comprehensive review is essential to capture the current landscape and future directions for machine learning in depression prediction.

The primary aim of this paper is to systematically review

TABLE I
THE TAXONOMY OF DEPRESSION PREDICTION WITH DATA MODALITY

Modality		Example	Feature	Strengths	Limitations	Applicability
Electrophysiological Signals	EEG		· Alpha asymmetry · Theta wave abnormalities · Beta wave abnormalities	· Directly monitor brain and heart activity to capture subtle real-time changes in emotion and cognition	· Subject to the influences of environmental noise, equipment quality, and operator expertise · Some data collection methods are semi-invasive, which may affect comfort and interfere with physiological functions	· Particularly well-suited for real-time physiological monitoring and depression prediction, especially in the contexts of sleep and cognitive research
	HRV /ECG		· Increased or decreased heart rate · Decreased HRV · Prolonged QT interval			
Brain Imaging	MRI		· Hippocampal atrophy · Prefrontal cortex abnormalities · Connectivity abnormalities	· High spatial resolution, allowing precise insights into brain structure and function	· Expensive, requires complex equipment and specialized operation · Data interpretation demands deep understanding of neural pathways	· Best for in-depth analysis of brain regions, studying the neurobiological mechanisms of depression, and longitudinal assessments of treatment effects
	fNIRS		· Reduced activity in the prefrontal corte · Increased activity in the amygdala · Basal ganglia dysfunction			
Audiovisual Data	Facial expression		· Reduced facial expressions · Down-turned mouth corners · Tense muscles · Drooping eyelids	· Non-invasive, useful for assessing emotional and social dysfunctions · Cost-effective and accessible for large-scale screenings	· Influenced by individual differences and environmental conditions · Having different meanings across cultures	· Suitable for large-scale screenings and identifying emotional and social symptoms, especially in cost-sensitive applications
	Speech		· Slowed speech · Lower volume · Monotonous pitch · Increased silence and pauses			
	Gait		· Slower walking pace · Reduced step length · An unsteady gait · Increased body sway			
Text			· Posting negative content · Decreased interaction · Irregular posting times · Singular themes	· Non-invasive and easily accessible; · Offering insights into subjective experiences, emotions, and cognition	· Subjective nature and diverse sources require extensive preprocessing · Social media may be influenced by self-presentation biases	· Ideal for long-term monitoring and large-scale analysis of subjective experiences, particularly in social and cognitive research

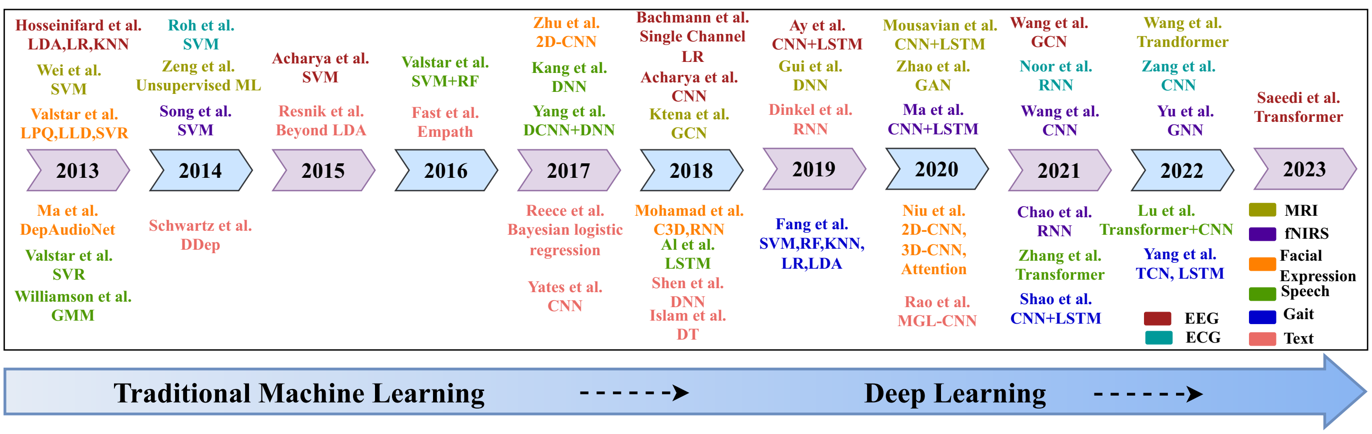


Fig. 1. The evolution of depression prediction methods.

and analyze the application of machine learning in depression prediction, highlighting recent advancements across various data modalities and revealing future research directions and challenges. Specifically, we focus on two main approaches: traditional machine learning and deep learning. We provide an overview of the research process in machine learning, summarize typical methods, and offer a comparative analysis of the strengths and limitations of each approach. Furthermore, we explore future directions in light of recent advancements in intelligent computing and big data technologies. In the application of machine learning, the choice of data modality is crucial for both model performance and interpretability. By analyzing the different data modalities, we delve into their potential and challenges in depression prediction, thereby providing strong support for methodological advancements. As shown in Table I, we focus on methodological advancements and potential research directions in four main data modalities: ***electrophysiological signals***, ***brain imaging***, ***audiovisual data***, ***text***, and ***mixed data***.

(1) *Electrophysiological signals* are changes in the membrane potential of individual cells recorded by surface metal electrodes [55]. Electroencephalography (EEG) and electrocardiography (ECG)/heart rate variability (HRV) are clinically relevant electrophysiological signals that are widely employed in depression prediction due to their noninvasive detection [56]–[61].

(2) *Brain imaging* refers to noninvasive or minimally invasive techniques that enable imaging of the brain's structure or function [62], Magnetic resonance imaging (MRI), functional near-infrared spectroscopy (fNIRS), etc., are its primary data representations. Research suggests that brain imaging can identify different types of depression based on the affected

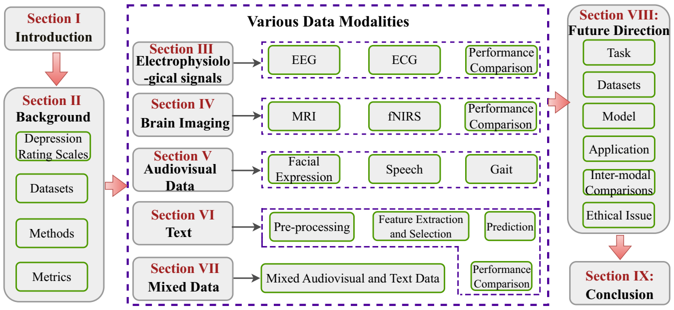


Fig. 2. The schematic structure of this paper and the relationship between the adjacent sections. The main body of this survey focuses on various data modalities for depression prediction, including electrophysiological signals, brain imaging, audiovisual data, and text. For each modality, we provide a comprehensive overview of the machine learning pipeline, encompassing data preprocessing, feature extraction and selection, prediction, and performance comparison. Finally, based on a synthesis of the characteristics and research findings for each data modality, we propose directions for future.

brain region [63]–[68].

(3) *Audiovisual data*, including facial expressions, speech, and gait, are captured by video and audio recording devices. Slower movement and speech are common clinical manifestations of depression [69], which can be audio-visually captured and used as a valuable tool in depression prediction [70]–[73].

(4) *Text* is a direct medium for expressing thoughts and emotions. Social media platforms provide an avenue for people to express their true feelings. Numerous studies have examined textual expression disparities between depressed and nondepressed individuals in social media settings [46], [74]–[80].

(5) *Mixed data* provide more features than single modality data, resulting in a relatively comprehensive understanding of the symptomatology of depressed patients [81]–[84].

This analysis encompasses over 200 research contributions, providing a systematic and in-depth survey of depression prediction. The primary contributions of our work, compared to existing literature, are as follows:

(1) We present a comprehensive overview of methodologies across diverse data modalities in depression prediction, including electrophysiological signals, brain imaging, audiovisual data, and text. A single data modality may not fully capture the complexity and multidimensional characteristics of depression. Through a detailed comparative analysis of each modality's strengths and limitations, we offer diverse insights that enhance understanding and predictive accuracy for depression.

(2) We outline the machine learning research process within each modality, summarizing prevalent methodologies and providing a comparative analysis of the advantages and limitations of various approaches. This serves as a valuable reference for model selection and optimization in depression prediction across different modalities.

(3) We delve into future directions and potential challenges in depression prediction across several dimensions—including task types, datasets, model, applications, inter-modal comparisons, and ethical issue a comprehensive and insightful research survey to guide future studies.

Our survey review focused on research conducted in the past decade, from 2013 to 2023. Several of the significant methods in the field of depression prediction developed during this period are shown in Fig 1. This time frame was chosen because of the rapid developments in machine learning in recent years, and we aimed to reflect the latest advancements and trends in this field. Finally, the papers included in our survey had to meet the following criteria: (1) Explicitly discussing the prediction of depression. (2) Applying at least one machine learning technique. (3) Publishing in peer-reviewed journals or conference proceedings. (4) Providing sufficient methodological details for the assessment of research quality.

The remainder of this survey is structured as follows: Section II introduces fundamental concepts related to depression prediction tasks, including scales, datasets, and methods. Sections III to VII provide an in-depth analysis of machine learning-based approaches to depression prediction, focusing on various data modalities. In Section VIII, we explore potential areas for future research. Section IX concludes the survey. In addition, for the sake of clarity, the structure of this paper is graphically illustrated in Fig. 2.

II. Background

This section provides an overview of depression prediction tasks, including scales, datasets, and methods.

First, depression scales utilized as predictive labels are introduced. Second, some publicly and privately datasets are presented. Third, the process of developing depression prediction models based on various modalities is outlined.

A. Depression Rating Scales

Depression rating scales, commonly employed in psychiatric practice, provide a straightforward means of capturing depressive symptoms, reflecting patients' subjective experiences. While serving as predictive tools for depression, final diagnoses rely on clinical judgment and medical data analysis. However, due to privacy considerations, patient information is typically safeguarded from public access, leading most studies on depression prediction to rely on questionnaire responses as the primary data source. (Due to length limitations for regular submissions, we provide the standard depression test scales in the supplementary material).

B. Datasets

This section introduces publicly available datasets with name, subject, and data modality information, as shown in Table II. Due to privacy concerns, only a few data modalities have publicly available datasets. Audio, video and text data are easier to collect, so most publicly available datasets are related to these modalities. Due to space limitations, here we only briefly show some representative and frequently used public datasets. Additionally, we present privately managed datasets that are used only by researchers for their analyses.

The Multimodal Open Dataset for Mental-disorder Analysis (MODMA) [86] comprises EEG data and audio recordings, mainly from clinically depressed patients and their matching normal controls, serving as a multimodal open dataset for mental disorder analysis.

The Audio/Visual Emotion Challenge (AVEC) [25], [35], [88]–[90], [97], [98] is an expression recognition challenge

TABLE II
COMMONLY EMPLOYED DATASETS

Dataset	Modality	Subjects	Type	Public/Private
Acharya et al. [85]	EEG	15D+15C	Depression	Private
MODMA [86]	EEG	24D+29C	Mental-disorder	Public
	A	23D+29C		
Song et al. [87]	fNIRS	108D+30C	Depression	Private
AVEC2013 [35]	A+V(Facial images)	292	Depression	Public
AVEC2014 [88]	A+V (Facial images)	84	Depression	Public
DAIC-WOZ [25], [89]	A+V (Facial features)	60D+133C	Depression	Public
E-DAIC [90]	A+V (Facial images)	275	Depression	Public
EATD [91]	A+T (Transcripts)	30D+132C	Depression	Public
RSDD [92]	T	9210D+107274C	Depression	Public
eRisk [93]	T	135D+752C	Depression	Public
CLPsych 2015 [94]	T	327D+573C	Depression	Public
Bell Let's Talk [95]	T	53D+101C	Depression	Public
Yuan et al. [96]	Gait	54D+47C	Depression	Private

1 The subjects are divided into two types: people with depression (D) and healthy controls (C).
2 The data modalities are represented as audio (A), text (T) and video (V). Some datasets only provide depression scores for regression tasks and does not provide the number of depression and normal controls.

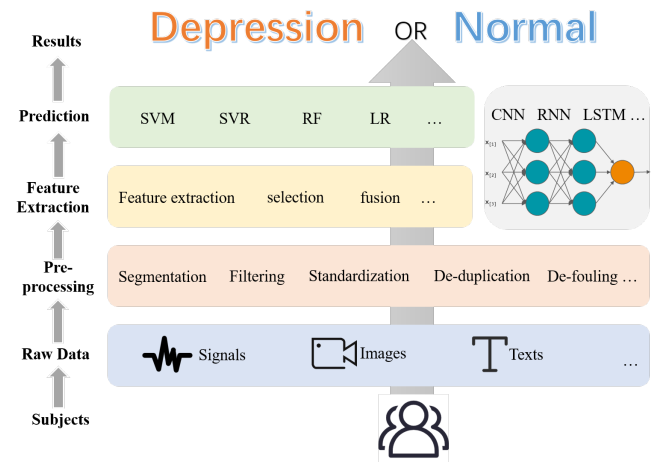


Fig. 3. Multiple-modality depression prediction methods.

that has been held every year since 2011. It is recognized as the top international competition in the field of emotional computing. AVEC2013 began to introduce the task of depression prediction, which considers the analysis of depression based on auditory vision as a classification or regression problem.

The Audio-Visual Depression Corpus (AViD-Corpus) [35], [88] contains 340 video clips of subjects performing a human-computer interaction task while being recorded by a webcam and a microphone. The speakers were recorded between one and four times, with a period of two weeks between the measurements. These data were used for the AVEC2013 and AVEC2014 Challenge, and AVEC2014 uses a subset of the AVEC2013 audio-visual depression corpus.

The Distress Analysis Interview Corpus/Wizard-of-Oz (DAIC-WOZ) [99] contains clinical interviews designed to support the diagnosis of psychological distress conditions such as anxiety, depression, and post-traumatic stress disorder. The collected data include audio and facial features and extensive questionnaire responses. These data were used for the AVEC2016 and AVEC2017 Challenge.

The Extended DAIC Database (E-DAIC) [90] is the extended version of the DAIC-WOZ database for depression and post-traumatic stress disorder assessment, developed by the Institute for Creative Technologies. These data were used for the AVEC2019 Challenge.

The Emotional Audio-Textual Depression Corpus (EATD-Corpus) [91] contains audio and extracted transcripts of responses from 162 depressed and nondepressed volunteers. EATD-Corpus is the first and only public depression dataset that includes audio and text data in Chinese.

The Reddit Self-reported Depression Diagnosis (RSDD) dataset [92] consists of Reddit posts for 9,210 users who have claimed to have been diagnosed with depression and 107,274 healthy users, which is a large-scale general forum dataset.

The early risk (eRisk) dataset [93] comprises text samples aggregated from postings by 887 Reddit users. The task's idea is to classify users into depression risk cases and non-risk cases.

The Computational Linguistics and Clinical Psychology (CL-Psych) [94] offers direct comparisons of methods for analyzing language related to mental health on social media. The dataset utilized includes tweets from users who state a diagnoses of depression or post-traumatic stress disorder.

Bell Let's Talk [95] is a campaign created by Bell Canada to help reduce stigma and promote awareness and understanding of mental health issues. Canadians started the dialog on mental health, contributing more than 122 million tweets, texts, calls and social media shares on Bell Let's Talk Day.

Furthermore, there are other public datasets that have been less utilized compared to those mentioned in the TableII. Due to space constraints, we have not provided detailed descriptions of these datasets here, such as Multiface [100], Early Mental Health Uncovering [101], Moodable and EMU datasets [76], Black Dog Institute [102] and University of Pittsburgh depression dataset [103].

C. Methods

Although a range of methods have been proposed for depression prediction based on different modality, the underlying predictive procedure remains consistent, as shown in Fig 3.

1) Data: Depression prediction relies on two main types of data: clinical data (e.g., EEG, fMRI) collected via clinical sensors, providing objective physiological measurements, but requiring expensive equipment and expert interpretation; and daily data (e.g., audio-video recordings, social media data), offering easy access and authentic reflections of daily life but facing challenges like unstable quality and environmental noise interference during acquisition and processing [104].

2) Pre-processing: Raw data often contain significant noise from the capture device and environment. For instance, an electrophysiological signal may be affected by baseline drift, power-line interference, and electromyography signals caused by equipment and the human body [105]–[107]. Audiovisual data are susceptible to environmental influences, resulting in considerable noise, such as background noise, variations in lighting, and motion blur. Thus, preprocessing raw data is necessary, and appropriate methods should be used to remove different noises. Preprocessing also allows data normalization and conversion [108], [109] to a proper format for subsequent feature extraction, matching the model's requirements and improving performance.

3) Feature extraction and selection: Feature extraction entails converting data to numerical attributes that preserve crucial information for depression prediction while enabling model differentiation. This process can involve either hand-crafted or data-driven approaches. Hand-crafted feature extraction involves the manual identification of attributes based on data properties and task requirements [110]–[112]. Data-driven techniques employ sophisticated algorithms, such as deep learning methods, to automatically extract features without human involvement.

Extracting diverse features from datasets often introduces a substantial amount of irrelevant and redundant information. This typically prolongs feature analysis and model training, increases the risk of overfitting, and diminishes the model's generalization capacity, ultimately leading to reduced recognition rates. Furthermore, feature sparsity can exacerbate model performance degradation. In such instances, employing an appropriate feature selection method can markedly enhance classification accuracy by isolating the most informative features.

4) Prediction: In depression prediction models, this task can be represented through a unified mathematical formulation. Let the input data be denoted by X and the predicted output by y. Regardless of whether traditional machine learning or deep learning methods are employed, the core objective of the model is to construct a mapping function $f(\cdot)$ that satisfies the following relationship:

$$y = f(X; \theta, \phi) \tag{1}$$

where X represents the input data, which may include various data modalities such as EEG, images, audio, or text. y denotes the output of the depression prediction model, typically expressed as a classification label or risk score. $f(\cdot)$ is the depression prediction model, defining the mapping from input X to output y. θ and ϕ are parameter sets corresponding to the feature extraction and prediction components, respectively.

In traditional machine learning methods, the feature extraction process $g(\cdot)$ is typically hand-crafted and does not rely on learning parameters. These manually designed features are then fed into classifiers or regression models for depression prediction [35], [85], [113]. The corresponding mathematical formula can be expressed as:

$$y = h(g(X); \phi) \tag{2}$$

where $g(X)$ is the hand-engineered feature function, which transforms the original data X into the feature space. $h(\cdot)$ is the classifier or regression model parameterized by ϕ, which may include models such as Support Vector Machine (SVM) [114], Support Vector Regression (SVR) [115], Random Forest (RF) [116], and Logistic Regression (LR) [117]. Deep learning methods typically adopt an end-to-end structure, enabling automatic feature extraction and depression prediction within a single learning process [105], [118]–[120]. As shown in 1, feature extraction and prediction are integrated into a unified model, with parameters θ and ϕ being jointly learned. This approach is applicable to various deep learning architectures, such as Convolutional Neural Networks (CNNs) [121], Recurrent Neural Networks (RNNs) [122], Long Short-Term Memory networks (LSTMs) [123], and Transformers [124].

In the subsequent sections, we outline the plan based on specific modality data, covering data preprocessing, feature extraction and selection, and prediction. Feature extraction and selection are mainly targeted at traditional machine learning methods that require manual feature extraction, as deep learning methods typically operate in an end-to-end manner without the need for additional feature processing. Additionally, we summarize the experimental results for each modality to facilitate comparative discussions.

5) Parameter Tuning and Optimization: In machine learning based prediction of depressive disorders, effective parameter tuning and optimization strategies are crucial for enhancing model performance, especially when working with diverse data modalities (e.g., text, audio, EEG signals). To guide future researchers in effective parameter tuning, we recommend considering both traditional and model-free optimization methods. **(1) Traditional Optimization Methods.** It primarily include grid search, random search, evolutionary algorithms, adaptive tuning strategies, and combined tuning approaches [125]–[130]. These methods rely on specific search processes or optimization rules to ensure that the optimal parameters are identified within a defined tuning space. For example, a combination of grid search and random search can be applied to models using speech and text data: random search initially identifies a broad parameter range, followed by grid search for fine-tuning, optimizing the extraction of audio and text features. Adaptive tuning strategies are valuable for handling physiological signals with high individual variability (such as EEG), automatically adjusting parameters to improve the model's robustness.In deep learning, due to the high dimensionality and non-linearity of models, parameter optimization becomes more complex. Therefore, selecting the appropriate optimization algorithm is crucial for improving both training efficiency and model performance. Gradient descent methods are widely used in deep learning, with adaptive optimization algorithms such as Adam and Momentum being particularly effective in accelerating convergence and enhancing model performance in tasks [131]–[133]. **(2) Model-Free Optimization Methods.** It enable optimization without relying on gradient information, making them especially suitable for non-continuous or complex loss functions. As a model-free optimization technique, the Beetle Antennae Search (BAS) algorithm efficiently optimizes the non-continuous or complex loss functions commonly encountered in multimodal depressive disorder recognition tasks [134], [135]. Further, advanced BAS variants, such as fallback beetle antennae search algorithm, Beetle Antennae search (BAS) algorithm, called BAS-adaptive moment estimation [136], [137], can offer robust global search capabilities, low time complexity, dynamic adaptability, and enhanced robustness, making it highly effective for supporting multimodal data integration and optimization in depressive disorder recognition tasks. Future research can explore model optimization based on BAS and its improved variants to enhance model performance and generalization. **(3) Hyperparameter optimization.** Hyperparameter optimization plays a pivotal role in the performance of deep learning models, particularly for complex tasks such as depression prediction. Typically, researchers rely on experience and domain knowledge to select initial hyperparameter values, guided by factors such as model type, data characteristics, and task requirements. Validation set testing remains a common approach for optimization, where various hyperparameter configurations are evaluated to identify the combination that maximizes model performance. However, this process is resource-intensive, time-consuming, and task-specific, as optimal hyperparameters often vary across tasks, making it challenging to standardize.

III. Electrophysiological signals

Depression, a mental disorder characterized by severe mood fluctuations and aberrant brain activity, can be detected by EEG-monitored brain activity [138], [139]. Depressive episodes may also present with symptoms such as nausea, vomiting, chest tightness, and sweating, indicative of autonomic dysfunction [140]. Signals such as ECG, heart sounds, and pulse can partially capture these alterations, highlighting the electrophysiological differences between depressed and healthy individuals. Consequently, depression prediction using these signals has emerged as a prominent research area and a promising direction. Given the infrequent use of other signal types, we focus on EEG- and ECG-based depression prediction. Additionally, due to the scarcity of relevant regression tasks, our attention is centered on classification tasks.

A. EEG

EEG, a direct and objective representation of human neurological brain activity, closely correlates with brain function and emotional state. EEG provides a real-time depiction of emotional changes and high temporal resolution for examining brain dynamics [141]–[143]. Consequently, EEG is a promising signal for depression prediction [144]. Current research predominantly employs conventional machine learning methods [145], while some scholars explore deep learning algorithms, resulting in significant advancements [47].

1) Pre-processing: EEG device-captured signals are low-resolution due to poor signal-to-noise ratios, impaired by physiological artifacts such as muscle movements, eye blinks, and heartbeats, as well as non-physiological factors including electrode placement, ambient noise, and device faults. For accurately identify depression, it is essential to remove noise and artifacts from raw data. Artifact removal can be automated or manual. Automated methods typically involve filtering [146]–[148] and Independent Component Analysis [148]–[150], which, however, may alter the inherent dynamics of the signal. In some studies, the task of artifact removal is manually carried out by experts with extensive experience [22], [151]. They meticulously examine each signal to pinpoint the exact locations of artifacts, selecting periods devoid of artifacts for analysis to minimize any changes to the signal being studied.

Moreover, deep learning preprocessing also involves min-max normalization [149], [151], [152], converting 1D EEG signals into 2D image matrix [150], and the creation of functional connectivity maps [153]. These steps enrich the data representation, providing a higher dimensionality of information that allows deep-learning models to detect and learn complex, depression-related patterns more effectively.

2) Feature extraction and selection: Features derived from EEG signals can be broadly categorized into linear analyses and nonlinear analyses. For EEG-based auxiliary depression diagnosis, linear analysis encompasses time and frequency domain analyses. Commonly employed time domain features include the mean, variance, kurtosis, and skewness of the EEG signal, which vary with the EEG signal's sliding window and accurately represent its temporal evolution. Frequency domain analysis involves characterizing nonstationary EEG signals based on their frequency components and calculating essential properties at each frequency. Nonlinear approaches capture the chaotic behavior and abrupt changes in EEG signals resulting from brain-based physiological phenomena [154]. In EEG-based auxiliary depression diagnosis research, notable nonlinear features encompass correlation dimension [155], the Lyapunov exponent [156]–[158], entropy [159]–[161], Higuchi's fractal dimension, relative wavelet energy, and detrended fluctuation analysis, among others. Each nonlinear indicator reflects distinct aspects of EEG, uncovering information that other methods might overlook.

Different methods have been explored for the purpose of feature selection and dimensionality reduction for an application of EEG based diagnosis for depression. Among them, the most popular are Principal Component Analysis [147], [148], [160], Genetic Algorithm [162], [163], and t-test [164].

3) Prediction: The prediction of depression using traditional machine learning methods often involves extracting diverse features from EEG and combining them with classifiers such as SVM [33], [85], [146], [165], LR [144], [160], [166], and linear discriminant analysis (LDA) [148]. This category of methodologies provides a dependable theoretical structure for accurately forecasting depression. Importantly, they provide a high level of interpretability when analyzing the expected outcomes.

Deep models employed in EEG-based depression prediction studies are classified into three primary categories: (1) Conventional neural networks, such as Multilayer Perceptrons and Probabilistic Neural Networks [164], rely on manually designed features as input. (2) End-to-end architectures, such as CNNs [22], [39], [151], [167], LSTMs [153], and transformers [120], [168], can automatically extract features from EEG data. (3) Architectures like CNNs or GCNs extract critical spatio-temporal features from high-level data representations [150], [152], [169]–[171]. These representations include feature matrices, functional connectivity networks, or 2D images, which are transformed based on various feature extraction techniques.

B. ECG

HRV refers to the physiological variation in the interval between consecutive heartbeats [172]. Traditionally, HRV analysis is performed on ECG signals and has been established as a helpful tool for depression prediction [173]. Studies show that individuals with psychiatric disorders, such as depression, exhibit reduced HRV, indicative of reduced parasympathetic control and increased sympathetic activity [174]. Recent research on HRV-based depression prediction identifies two primary strategies: traditional methods using hand-crafted features and end-to-end deep learning methods. (Due to length constraints, we provide a detailed description of the technique in the supplementary material.)

TABLE III
EXPERIMENTAL RESULTS BASED ON ELECTROPHYSIOLOGICAL SIGNALS. WE SUMMARIZE REPRESENTATIVE MACHINE LEARNING METHODS AND REPORT THE EVALUATION METRICS: ACCURACY, SENSITIVITY, AND SPECIFICITY.

Type	Paper	Dataset	Methods	Accuracy	Sensitivity	Specificity
EEG	Acharya et al. [85]	15D+15C	SVM	98.00	97.00	98.85
	Mumtaz et al. [165]	34D+30C	SVM	98.00	99.90	95.00
	Liao et al. [33]	12D+12C	SVM	80.00	83.33	83.33
	Sharma et al. [146]	15D+15C	LS-SVM	99.54	98.66	99.38
	Bachmnn et al. [166]	13D+13C	LR	92.00	-	-
	Hosseinifard et al. [144]	45D+45C	LR	90.00	-	-
	Cukic et al. [160]	23D+20C	LR	97.56	-	-
	Cai et al. [147]	152D+113C	DT	76.40	-	-
	Mahato et al. [148]	30D+30C	LDA	93.33	86.67	88.24
	Faust et al. [164]	30D+30C	PNN	99.50	-	-
	Wan et al. [167]	23D+12C	HybridEEGNet	79.08	68.78	84.45
	Saeedi et al. [153]	34D+30C	1DCNN-LSTM	99.24	98.52	100.00
	Thoduparambil et al. [149]	-	CNN-LSTM	99.07	99.50	98.60
	Acharya et al. [22]	15D+15C	CNN	95.96	94.99	96.00
	Kang et al. [151]	34D+30C	CNN	98.85	98.84	98.66
	Sharma et al. [39]	21D+24C	DepHNN	99.10	-	-
	Li et al. [150]	24D+24C	ConvNet	85.62	-	-
	Ay et al. [175]	15D+15C	CNN-LSTM	99.12	99.11	-
	Wang et al. [152]	MODMA	GCN	92.23	98.32	85.96
	Wu et al. [171]	MODMA	GCN	93.85	-	-
	Sun et al. [169]	MDD	GCN	99.29	99.37	99.32
	Wu et al. [170]	MODMA	GCN	99.03	-	-
	Saeedi et al. [168]	34D+30C	Transformer	97.22	99.41	94.87
	Tao et al. [120]	28D+37C	Transformer	92.80	-	-
ECG	Byuna et al. [176]	31D+41C	SVM	74.40	73.00	75.60
	Roh et al. [177]	23D	SVM	71.00	-	-
	Matsui et al. [178]	13D+28C	LDA	88.00	85.00	89.00
	Sun et al. [179]	44D+47C	LR	79.00	80.00	79.00
	Kuang et al. [180]	38D+38C	Bayesian	86.40	89.50	84.20
	Noor et al. [181]	5000D	RNN	97.24	-	-
	zang et al. [182]	37D+37C	CNN	93.96	89.43	98.49
	Mohanraj et al. [183]	15D+15C	DesNN	90.00	-	-

[1] People with depression (D) and healthy controls (C).

C. Performance Comparison

Table III summarizes the experimental results for depression prediction using electrophysiological signals, including studies based on EEG and ECG data. The following key observations can be drawn from the table:

(1) The classification accuracy for depression prediction using EEG signals ranges from 76.40% to 99.54%, with sensitivity ranging from 68.78% to 99.99% and specificity from 83.33% to 100%. These performance variations across studies may be influenced by factors such as sample size, feature extraction techniques, and the choice of machine learning models. For instance, Cai et al. [147] report the lowest accuracy of 76.40%, which could be attributed to the larger sample size in their dataset. Larger sample sizes often introduce greater variability and complexity, posing challenges for models to achieve high accuracy.

(2) The accuracy of depression prediction using ECG signals ranges from 71.00% to 97.24%. Noor et al. [181] achieve the highest accuracy of 97.24% with an RNN-based approach, demonstrating the effectiveness of recurrent neural networks in modeling time-series data. Compared to traditional methods, deep learning exhibits significantly better performance on ECG data, underscoring its potential for advanced depression prediction tasks.

(3) Table III presents studies utilizing various methods, including traditional machine learning techniques (e.g., SVM, LR, and LDA) and deep learning approaches (e.g., CNN, GCN, and transformer). Overall, deep learning methods demonstrate superior performance, particularly with large datasets. However, some studies, such as those by Wan et al. [167] (CNN, 79.08%) and Li et al. [150] (ConvNet, 85.62%), exhibit significant performance variability, likely due to cross-validation not providing the same performance gains as experiments with large sample sizes. Traditional methods, often supported by statistical significance analysis, show greater stability and robustness on smaller datasets. In contrast, some deep learning studies lack sufficient significance testing, which reduces the statistical reliability of their findings. Future research should incorporate significance analysis (e.g., t-tests or ANOVA) to enhance the robustness and reliability of deep learning model results.

IV. BRAIN IMAGING

Depression can induce hippocampal atrophy, alter neurotransmitters, and provoke chronic inflammation, resulting in memory impairment, slowed cognition, and fatigue. These symptoms significantly diminish patients' energy levels and motivation, hindering their ability to socialize [184], [185]. These findings provide a pivotal point for depression prediction. Brain imaging involves capturing structural and functional brain data using noninvasive, high-precision instruments that interact with brain tissue and various energy forms (e.g., electromagnetic or particle radiation). Brain imaging techniques include MRI, fNIRS, computed tomography, and positron emission tomography. These methods offer several benefits, such as safety, painlessness, and noninvasiveness, which are crucial for depression analysis. This review primarily focuses on MRI- and fNIRS-based prediction of depression.

A. MRI

MRI is a noninvasive imaging modality that generates detailed, three-dimensional anatomical images depicting changes in brain tissue volume, structure, and neural activity. Techniques include structural MRI (sMRI) and functional MRI (fMRI). sMRI provides high-resolution images of the brain's anatomy, objectively reflecting structural alterations within the brain [186]–[188]. fMRI assesses hemodynamic changes induced by neuronal activity, revealing the relationship between brain activity and cognition. Most depression-related MRI prediction research employs fMRI, which is primarily categorized into resting state fMRI and task state fMRI. Task state fMRI is model-driven, necessitating meticulous experimental design with model quality closely linked to experimental design. Conversely, resting state fMRI does not involve specific cognitive tasks, requiring participants to remain calm, relaxed, and alert [189].

1) Pre-processing: In fMRI-based depression prediction approaches, input data typically undergo pre-processing, which commonly involves smoothing and denoising the data using filter and Gaussian kernel [40], [190]. Additionally, pre-processing steps often include motion correction, slice-timing correction, and normalization to account for variations in scanner acquisition protocols and individual anatomical differences. These procedures aim to enhance the quality of the fMRI data and reduce potential confounding factors before further analysis. Moreover, nuisance signal regression techniques may be applied to remove physiological noise sources, such as cardiac and respiratory fluctuations, further refining the data for accurate depression prediction modeling.

2) Feature extraction and selection: In traditional fMRI-based machine learning methods, nonlinear features such as functional connectivity, degree centrality, regional homogeneity, and amplitude of low-frequency fluctuation are employed for analysis [191]. Additionally, there are manually extracted features, such as statistical features and channel correlations, which can serve as additional inputs to the model for comprehensive analysis [23]. Together, these measures offer a comprehensive understanding of brain function and connectivity alterations associated with mental disorders.

Moreover, many researchers opt for the least absolute shrinkage and selection operator (LASSO) to capture and optimize critical features. They employ group, sparse group, and standard LASSO for analysis [192], [193].

3) Prediction: Multivariate pattern analysis [194] is employed for individual-level prediction of mental disorders. This method determines whether an unknown sample exhibits specific diseases by modeling multivariate data of a predetermined type. Furthermore, other classical classification methods such as SVM are also widely used for depression prediction [34], [195]–[198].

Prediction models based on deep learning encompassing CNNs, LSTM, GCNs, transformers and capsule network [199]. These approaches leverage different aspects of the fMRI data, such as spatial correlations, temporal changes, functional connectivity patterns, and long-range dependencies, to extract informative features and improve prediction accuracy. While CNNs and LSTM networks focus on local spatial correlations and temporal changes [40], [200], respectively, GCNs and Transformers capture functional connectivity patterns and long-range dependencies [201]–[204]. Capsule Networks offer a promising alternative to CNNs, demonstrating superior generalization capabilities in complex image classification tasks [205].

B. fNIRS

fNIRS utilizes near-infrared light scattering in blood to measure changes in oxygenated hemoglobin, deoxygenated hemoglobin, and total hemoglobin during brain activity. During fNIRS data acquisition, a series of motor or cognitive tasks are typically administered to elicit brain responses [206], [207]. These changes reflect neuronal and cognitive brain function, making fNIRS a valuable tool for depression prediction [208]–[210]. (For detailed method description, refer to supplementary material due to submission length constraints.)

C. Performance Comparison

TABLE IV
EXPERIMENTAL RESULTS BASED ON BRAIN IMAGING. WE SUMMARIZE REPRESENTATIVE MACHINE LEARNING METHODS AND REPORT THE EVALUATION METRICS: ACCURACY, SENSITIVITY, AND SPECIFICITY.

Type	Paper	Dataset	Methods	Accuracy	Sensitivity	Specificity
fMRI	Gu et al. [211]	46D+46C	Multivariate pattern analysis	90.22	-	-
	Yamashita et al. [193]	660PD+924C	LASSO	66.0	72.00	65.00
	Zeng et al. [212]	24D+29C	Unsupervised ML	92.50	-	-
	Lord et al. [198]	21D+22C	SVM	99.31	-	-
	Wei et al. [34]	20D+20C	SVM	90.00	-	-
	Yu et al. [192]	31D+31C	SVM	94.68	94.48	94.87
	Bhaumik [197]	38D+29C	SVM	76.10	81.50	68.90
	Wang et al. [196]	31D+29C	SVM	95.00	96.77	93.10
	Yan et al. [195]	43D+56C	Dynamic functional connectivity+SVM	95.59	97.77	94.68
	Mousavian et al. [213]	38D+241C	Cubes-Atlas	93.00	82.00	95.00
	Gui et al. [200]	19D+20C	DNN	94.68	-	-
	Mousavian et al. [40]	38D+241C	CNN	100.00	86.00	0
	Zhao et al. [190]	269D+286C	Functional network connectivity +GAN	70.10	73.50	66.50
	Wang et al. [204]	282D+251C	Functional connectivity network+Graph neural networks	63.60	67.30	59.30
	Jun et al. [203]	29D+44C	GCN	74.10	56.60	86.90
	Cheng et al. [205]	70D+30C	Capsule Network	88.79	-	-
fNIRS	Song et al. [87]	108D+30C	SVM	86.77	-	-
	Song et al. [214]	108D+30C	SVM	89.71	-	-
	Zhu et al. [215]	10D+10C	One Rule	85.00	90.00	80.00
	Zhu et al. [207]	14D+17C	RF+eXtreme Gradient Boosting	92.60	84.80	91.70
	Ma et al. [206]	36D+48BD	LSTM	96.20	-	-
	Yu et al. [216]	79D+17C	Graph neural networks	87.50	-	-
	Chao et al. [217]	16D+16C	Cascade forward neural network	99.94	-	-
	Wang et al. [23]	79D+17C	CNN	90.00	-	-

[1] People with depression (D), psychiatric disorders (PD) Bipolar Disorder (BD) and healthy controls (C).

Table IV presents the performance of brain imaging depression prediction approaches. By analyzing these approaches on identical datasets, we derive three key observations:

(1) The general linear model is the primary analysis method for fMRI images. Its superior feature extraction capabilities have outperformed traditional linear feature extraction methods in fNIRS.

(2) When analyzing fMRI, CNN-based models demonstrate a stronger capacity to identify depression characteristics than functional connectivity network models. Deep learning methods have uncovered spatial and temporal relationships between two voxels in stationary state networks. CNNs are recognized as the most effective method for extracting spatial features.

(3) The need for more training samples poses a significant challenge when applying deep learning methods to medical image analysis. Although generative adversarial network (GAN)-based models have not yielded optimal results, this framework is promising for broad applicability and has substantial potential for neuroimaging biomarker identification.

V. AUDIOVISUAL DATA

Recently, researchers have increasingly explored the use of behavioral signals for depression prediction. They analyze abnormal expressive behaviors associated with depression, such as sluggish facial expressions, frequent avoidance of eye contact, speaking in short, flat-toned sentences, and exhibiting a disheveled posture. Depression prediction based on behavioral signals typically involves video (facial expression and gait) and audio (speech).

A. Facial expression

The face can discern a person's personality and mood. Some studies have identified differences in appearance and temperament between patients with depression and healthy individuals. Zhu et al.'s [218] research team from the Institute of Psychology of the Chinese Academy of Sciences analyzed 100 individuals with mental illnesses. They observed that these patients appeared sad, often frowning while reading aloud, with drooping mouth corners and tears in their eyes. Thus, comparing and analyzing facial expression data from depressed and healthy individuals can predict depression [219]–[224].

1) Pre-processing: The original facial video data typically contains a lot of redundant information, requiring manual processing to extract facial videos or images containing rich information. Subsequently, the data are annotated and subjected to face detection to determine the position of landmarks [225].

2) Feature extraction and selection: Depression prediction based on facial expression relies heavily on feature extraction, with researchers exploring various techniques to derive discriminative features. Key methods include the Motion History Histogram [226], Active Appearance Model (AAM) [225], [227], [228], local binary patterns (LBP) [229], and its variants [88], [230], [231]. The AAM captures dynamic facial expression features by combining shape and texture information, while LBP simplifies texture description without manual annotation, making it widely used in facial-based depression prediction. Improvements on LBP include the use of Local Gabor Binary Pattern for texture analysis [88] and the incorporation of curvelet transform [232] for extracting curvature information. Additionally, methods like space-time interest points, pyramid histogram of oriented gradients [233], perception-driven distance and area features [234] have been explored for facial feature extraction.

With the maturity of facial recognition technology, some open-source tools have been developed, such as Openface [235], Opensmile [236] and Facet [237], etc. These tools can extract various features such as facial action units, landmarks and eye gaze, providing many conveniences for analyzing faces.

Common feature selection methods for depression prediction based on facial analysis include principal component analysis [238], gaussian mixture model, and t-test [225]. Principal component analysis reduces the dimensionality of the feature space by linear transformation, retaining the most important feature information. Gaussian mixture model models the probability distribution of each feature to identify the most discriminative features. Meanwhile, t-test calculates the differences in features between different groups to select significant features for depression prediction.

3) Prediction: The most commonly used machine learning methods for depression prediction are SVM and its regression counterpart, SVR. Additionally, other regression techniques such as LR have also shown promising results.

In the realm of depression prediction based on facial expressions, deep learning techniques, including CNN and RNN, have become prominent. CNNs excel in spatial feature extraction [90], effectively capturing facial features and their relationships with depression levels [71], [239]. To overcome challenges posed by unsuitable poses, memory attention mechanisms are integrated to emphasize informative frames [240]. Further, temporal information combined with spatial information can reveal facial behavioral dynamics, such as slow head movements and eye contact avoidance in depressed individuals [241], [242]. Moreover, researchers have integrated dynamic historical histograms and two-channel/two-stream models [242], [243] with CNNs to capture spatiotemporal information [244].

However, solely relying on spatial and temporal information may hinder spatiotemporal relationship modeling, prompting the exploration of spatiotemporal correlations using 3D-CNNs [24], [245]. In addition, RNNs introduce a temporal dimension, enabling the modeling of time-varying attributes in facial video data [238]. Moreover, deep belief network [246] is a generative model that utilizes multiple layers of feature-detecting neurons. Different deep belief network models can combine 2D appearance features and 3D dynamic features of facial images [231]. Hybrid CNN-RNN models have shown promise in capturing both spatial and temporal features simultaneously, thus enhancing model performance in depression prediction tasks [247].

4) Performance Comparison: The performance of facial expression-based depression prediction methods is outlined in Table V. Comparing the performance of various approaches on identical datasets yields three primary observations:

(1) The performance of traditional machine learning methods is closely linked to manual feature extraction. As in facial expression recognition, local features such as the eyes and mouth in videos are crucial for video-based depression prediction. Combining these features with global features leads to improved results. Based on this, the work of Kaya et al. [248] achieved the best MAE (7.86) in the AVEC2013.

(2) In deep learning for depression prediction, CNNs are widely favored for their excellent spatial feature extraction. To address temporal variations in facial videos, researchers integrate RNNs or 3D convolutional kernels for capturing time-domain information, with 3D-CNN architectures showing enhanced performance. For example, the Melo et al. used the 3D-CNN method to achieve the best MAE of 5.96 and 5.82 in the AVEC2013 and 2014 datasets [243], [245].

(3) Deep learning methods outperform traditional machine learning techniques because they automatically learn high-level semantic features beneficial for depression prediction. 3D-CNN-based architectures better capture spatiotemporal facial information associated with depressive behavior in videos, enhancing model performance [245].

TABLE V
EXPERIMENTAL RESULTS BASED ON FACIAL EXPRESSION. WE SUMMARIZE REPRESENTATIVE MACHINE LEARNING METHODS AND REPORT THE EVALUATION METRICS: F1-SCORE, RMSE, AND MAE.

Paper	Dataset	Methods	F1-Score	RMSE	MAE
Valstar et al. [35]	AVEC2013	SVR	-	13.61	10.88
Valstar et al. [88]	AVEC2014	SVR	-	10.86	8.86
Alghowinem et al. [225]	30D+30C	SVM	71.20(Acc)	-	-
Alghowinem et al. [227]	30D+30C	SVM	75.00(Acc)	-	-
Kaya et al. [248]	AVEC2013	1-NN regressor	-	9.72	7.86
Jan et al. [229]	AVEC2014	LR	-	10.50	8.44
Wen et al. [249]	AVEC2013	SVR	-	10.27	8.22
Ringeval et al. [25], [89]	DAIC-WOZ	RF	58.30	6.97	6.12
Yu et al. [242]	AVEC2013	CNN	-	9.82	7.58
Yu et al. [242]	AVEC2014	CNN	-	9.55	7.47
Ringeval et al. [90]	E-DAIC	CNN	-	8.01	-
Melo et al. [24]	AVEC2013	CNN	-	8.26	6.40
Melo et al. [24]	AVEC2014	CNN	-	8.31	6.59
Ray et al. [250]	E-DAIC	LSTM	-	8.95	-
Melo et al. [239]	AVEC2013	CNN	-	8.25	6.30
Melo et al. [239]	AVEC2014	CNN	-	8.23	6.15
Zhou et al. [240]	AVEC2014	CNN	-	8.43	6.37
Melo et al. [245]	AVEC2013	3D-CNN	-	7.90	5.98
Melo et al. [245]	AVEC2014	3D-CNN	-	7.61	5.82
Zhou et al. [71]	AVEC2013	CNN	-	8.28	6.20
Zhou et al. [71]	AVEC2014	CNN	-	8.39	6.21
Melo et al. [243]	AVEC2013	3D-CNN	-	7.97	5.96
Melo et al. [243]	AVEC2014	3D-CNN	-	7.94	6.20
Wang et al. [251]	DAIC-WOZ	LSTM	78.30	-	-
Jazaery et al. [238]	AVEC2013	RNN-Convolutional 3D	-	9.28	7.37
Jazaery et al. [238]	AVEC2014	RNN-Convolutional 3D	-	9.20	7.22
Flores et al. [247]	DAIC-WOZ	CNN+LSTM	81.00	-	-

[1] People with depression (D) and healthy controls (C).
[2] Accuracy (Acc).

B. Speech

Clinical observations indicate that healthy and depressed individuals exhibit distinct speaking patterns on the phonetic scale. Depressed individuals typically speak more smoothly and monotonously, while healthy individuals speak more rhythmically, with fewer pauses and high-pitched response [224], [252]–[254]. Traditional speech-based depression analysis rely on these differences to design discriminative features. However, with the rapid advancement of deep learning techniques, speech-based depression prediction has shifted from hand-crafted acoustic features to deep learning-based frameworks [45], [255]–[258], addressing both depression classification and depression level regression. (Due to length constraints, we provide a detailed description of the technique in the supplementary material.)

C. Gait

Human gait is a daily movement that occurs in parallel with the development of higher brain structures and functions [259]. As a result, human gait reflects the integrity of the higher brain systems [260], rendering it a useful indicator of mental status. In addition, motion analysis of gait characteristics is conducted to evaluate depression patients [261]–[266]. Michalak [267] and Lemake et al. [268] observed reduced walking velocity, arm swing, vertical body movement, increased body sway, and more slumped posture in patients.

Compared to traditional biometrics utilized for detecting mental illness, gait is remotely observable, more difficult to imitate, and requires less cooperation from the subject [269], which makes it a promising source for depression prediction. The Kinect system [270], equipped with depth and infrared sensors, captures skeletal and RGB images during walking and is the primary acquisition device for gait-based depression prediction research. There are two basic approaches in related work: traditional machine learning and deep learning, which focus on the regression of scale scores and depression classification tasks. This section primarily presents research on the classification task, as the number of regression tasks is limited.

This study is based on two gait representations: skeleton and silhouette. The skeletal representation is derived from the 3D coordinate data of critical joints in the human body's structural model and characterizes human body behavior [271]. This representation is stable, less affected by the observational perspective, and accurately depicts variations in human action [272]. The silhouette representation is obtained by background subtraction followed by binarization. The gait energy image [273], produced by averaging the binary silhouette of the human body, is a popular silhouette feature due to its simple processing and strong anti-noise properties [274]. (Due to length constraints, we provide a detailed description of the technique in the supplementary material.)

VI. TEXT

Social networks offer a more convenient and accessible platform for people to communicate and express their emotions. Thus, researchers are analyzing data from social network users [46]. These studies demonstrated that individuals with depression exhibit distinct linguistic attributes and social behaviors when compared to nondepressed individuals [275]–[277]. Specifically, depressed individuals tend to use first-person pronouns, past tense verbs, and derogatory adjectives more frequently [278], [279]. Therefore, the analysis of social network data using computer technology provides a novel means to detect depression among users [280]–[283].

Social media is an excellent source of depression prediction using text. However, the data collection and annotation process differs from the other modalities. First, social media data is typically collected using large-scale crawlers and automated systems that extract information from multiple platforms. This contrasts with other data collection forms, where typically, a small number of users are recruited, and the data is collected interactively. Second, social media data exhibits diverse forms of language expression. Compared to other modalities, user-generated content on social media often tends to be more informal and unstructured and incorporates various modes of expression, including text, images, links, and emoticons. This diversity presents challenges in data annotation and analysis, as it requires considering the relationships and meanings across multiple modes of expression. Third, social media data is often noisy and contains much irrelevant information.

For example, individuals may post multiple photos or status updates, but this does not always translate into actual mental health issues. Domain experts may use various strategies to cleanse and select relevant data. In conclusion, there are significant differences in data collection and annotation processes between social media data and other modalities involving domain experts. Understanding and addressing these differences are key to conducting effective analysis and prediction using social media data.

Text-based depression analysis uses data from social media and online platforms, such as Twitter [36], [284], Reddit [277], [285], Weibo [284], [286], essays [279] and text messages [287]–[289]. However, privacy and ethical issues must be addressed, particularly with sensitive data like text messages, which is why most studies focus on public social platforms.

A. Pre-processing

The preprocessing methods for social text-based depression prediction tasks include steps such as removing interventions by virtual agents and non-verbal communication cues (transcribed text showing participant sighing or taking a deep breath and Emoji) [283], [290], tokenizing transcriptions into semantic units [283], normalizing features [287], [291], and segmenting words [292]. Among them, data cleaning and normalization techniques involve removing irrelevant information, such as non-English tweets, punctuation, and URL, etc. [291], [293].

B. Feature extraction and selection

Common methods for text feature extraction in social media include N-grams [294]–[296], emotion analysis [290], [297]–[299], bag of words [278], etc. By combining these characteristics, researchers can effectively analyze depressed users in social networks.

Additionally, the Linguistic Inquiry and Word Count tools are usually employed to extract features from social media [300]. These tools analyze the usage of different categories of words in texts to identify potential indicators of depression. Researchers often combine them with other features for analysis, such as effective features, mood tags, topics [276], personality traits [301] and N-grams [277]. Similarly, Empath [302] is a text analytics tool that utilizes a deep learning-based topic model to uncover topics in large-scale text data and comprehend the interrelationships between them. It is applicable for a multitude of tasks, including topic classification and sentiment analysis [287].

Moreover, researchers utilize statistical analysis methods for extracting features from text data to investigate depression. For instance, statistical measures like mean, variance, momentum, and entropy are applied to behavioral attributes such as social participation, emotion, linguistic style, and mention of antidepressants in social media posts [280]. Similarly, statistical features extracted from Instagram photos, including the number of comments and likes, presence of faces, and average hue, saturation, and value at the pixel level, are analyzed to understand depression-related patterns [303].

The feature selection methods often involve techniques such as Lasso [276] and principal component analysis [280]. These methods are mainly employed alongside supervised learning classifiers, such as SVM with radial-basis function kernels, and are evaluated using cross-validation techniques to ensure robustness and generalizability of the predictive models.

C. Prediction

Beyond traditional classifiers such as SVM and LR, topic modeling analyzes depression in social networks by predicting the topic distribution of documents. This method treats each document as a mixture of hidden topics, represented by the probability distribution of relevant terms. Latent dirichlet allocation [304] is a popular topic modeling method that provides a probability distribution of topics for each document, facilitating topic clustering or text classification [36], [305]. Many other methods have also achieved good results in depression prediction based on social networks, such as dictionary learning [286] and incremental classification [285], which provide more possibilities for depression analysis.

Deep learning methods, including DNNs, CNNs, RNNs, and transformer networks, have emerged as popular approaches for depression prediction using social text data. These methods leverage the inherent capacity of deep learning models to automatically extract and learn features from text data, facilitating the development of predictive models. Specifically, DNNs are employed for classification tasks, with adaptations such as momentum learning and cross-domain transformation to enhance model performance and adaptability across different data domains [284], [306]. CNNs are utilized for their ability to capture complex spatial features from text [27], [77], [92], while RNNs, especially LSTM and bidirectional LSTM models [292], [293], excel in processing sequential data, enabling the capture of long-term dependencies in text sequences. Transformer networks, with their attention mechanisms [283], offer enhanced model interpretability and performance improvements in depression prediction tasks by extracting hierarchical representations from text data.

D. Performance Comparison

Table VI illustrates the performance of text-based methods for depression prediction. By analyzing the results of various approaches on identical datasets, the following three observations are obtained:

(1) On the CLPsych 2015 dataset, compared with various traditional methods, the deep learning method CNN achieved the best F1-score 86.97 [307].

(2) On the RSDD dataset [92], the CNN model proposed by Yates et al. [92] uses a convolution layer to process posts to identify the features existing in the sliding text window. The CNN model proposed by Rao et al. [27] uses multiple gated leakage units to help the model capture context information. The results indicate that the sliding window approach outperforms the latter with an F1-score of 65.00, which suggests that for depression analysis, the essential information contained in the text is more crucial than the contextual information contained in the text.

TABLE VI
EXPERIMENTAL RESULTS BASED ON TEXT. WE SUMMARIZE REPRESENTATIVE MACHINE LEARNING METHODS AND REPORT THE EVALUATION METRICS: F1-SCORE AND ACCURACY.

Paper	Media	Dataset	Methods	F1-Score	Accuracy
Schwartz et al. [294]	Facebook	28749	Regression model	-	38.60
Nguyen et al. [276]	LiveJournal	38401D+229563C	LASSO	-	100.00
Pedersen et al. [295]	Twitter	CLPsych 2015	Lexical Decision List	74.20(Precision)	-
Resnik et al. [305]	Twitter	CLPsych 2015	Latent dirichlet allocation	62.00(Precision) 75.00(Recall)	-
Tung et al. [297]	Web post	18000	Event-driven depression tendency warning	62.40	-
Islam et al. [300]	Facebook	4149D+2996C	Decision Tree	73.00	-
Shen et al. [290]	Twitter	402D+36993C +300millionDC	Dictionary Learning	85.00	-
Reece et al. [303]	Instagram	71D+95C	RF	64.70	-
Wang et al. [298]	Micro-blog	122D+346C	Bayes	85.00	-
Nadeem et al. [278]	Twitter	CLPsych2015	LR	84.00	82.00
Tlachac et al. [287]	Text messages and twitter	Moodable and EMU	LR	80.60	77.30
Liu et al. [288]	Text message	219	LR	72.00(AUC)	-
Choudhury et al. [280]	Twitter	171D+305C	SVM	-	70.00
Tsugawa et al. [36]	Twitter	81D+128C	SVM	52.00	69.00
Wolohan et al. [277]	Reddit	4947D+7159C	SVM	81.80	72.90
Shatte et al. [299]	Reddit	365	SVM	67.00	66.00
Peng et al. [286]	Weibo	141D+135C	SVM	76.00	83.00
Li et al. [291]	Twitter	80million	SVM	-	86.98
Burdisso et al. [285]	Reddit	eRisk2017	Incremental classification	61.00	-
Mariñelarena et al. [306]	Reddit	83D+403C	DNN	82.93	94.81
Shen et al. [284]	Twitter, Weibo	1394D+1394C 580D+580C	DNN	79.00	-
Yates et al. [92]	Reddit	RSDD	CNN	65.00	-
Orabi et al. [307]	Twitter	CLPsych2015	CNN	86.97	87.96
Orabi et al. [307]	Facebook	Bell Let's Talk	CNN	82.25	83.12
Rao et al. [27]	Reddit	RSDD	CNN	54.00	-
Rao et al. [27]	Reddit	eRisk2017	CNN	60.00	-
Trotzek et al. [77]	Reddit	eRisk 2017	CNN+LR	73.00	-
Sadeque et al. [308]	Reddit	eRisk2017	SVM+RNN	64.00	-
Almeida et al. [296]	Reddit	eRisk 2017	Information Retrieval +Supervised Learning	53.00	-
Mallol-Ragolta et al. [283]	Interview transcript	DAIC-WoZ	Transformer	66.00(Recall)	-
Dinkel et al. [293]	Interview transcript	DAIC-WOZ	Bidirectional gated recurrent unit	84.00	-86.00
Hu et al. [292]	Weibo	28455D+28455C	Bi-LSTM	94.69	94.87

[1] People with depression (D) and healthy controls (C).

(3) The hybrid model outperforms the single models. On the eRisk2017 dataset [93], Trotzek et al. [77] achieved relatively good results (F1-score 73.00) by combining the CNN and LR models. Specifically, a CNN using various word embeddings and a logistic regression model based on user-level linguistic metadata are combined to improve detection performance.

VII. MIXED DATA

Depression prediction models using single modality data have made significant progress, but they face certain limitations. For instance, gait abnormalities can result from physiological disorders, obtaining physiological signals is challenging, and environmental impacts can affect voice and facial expressions. Therefore, a more robust depression recognition model can be developed by integrating complementary information from diverse modalities that capture distinct aspects of depression. Most mixed-data ways for analyzing depression rely on easily collectible data such as audio, video, and text. Thus, in this section, we focus on discussing depression prediction methods that utilize these modalities.

A. Mixed audiovisual and text data

1) Fusion methods: Multimodal data analysis involves integrating different types of data to predict depression, and

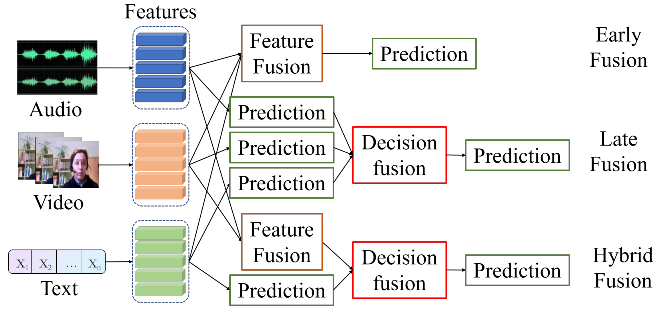


Fig. 4. Different approaches for fusion technique: Early fusion method, Late fusion method, and Hybrid fusion.

effective fusion of these modalities is a crucial step. Three fusion methods are commonly employed: early fusion, late fusion, and hybrid fusion, as shown in Fig 4.

Early fusion, integrates features extracted from each modality into a single feature vector for prediction. This method allows for the simultaneous processing of multiple modalities at the feature level, enabling joint learning and interaction between different types of data. The drawback of feature fusion lies in its susceptibility to the curse of dimensionality. Combining more modalities results in higher-dimensional feature vectors. To address this challenge, Principal component analysis is often employed to reduce the dimensionality of the combined features before prediction [309], [310].

In contrast, late fusion involves training separate models for each modality and combining their predictions at a later stage, allowing for more independent processing of individual modalities. Score Fusion and Decision Fusion are two late fusion methods commonly used in multimodal data analysis. In Score Fusion, scores from different modalities, such as probability estimates or likelihoods, are combined before making a classification decision. This method offers flexibility but may require careful weighting and can be sensitive to fusion techniques. On the other hand, Decision Fusion involves training multiple classifiers on different feature sets and aggregating their outputs to make a final decision. While Decision Fusion benefits from classifier diversity and can incorporate domain knowledge, it may be more complex and computationally intensive [309].

Finally, hybrid fusion combines the outputs of early fusion with the predictions from a single modality, leveraging the advantages of both approaches. Each fusion method has its unique strengths and may be suitable for different applications based on the characteristics of the data and the specific requirements of the prediction task.

2) Performance Comparison: The performance of multimodal depression prediction approaches is summarized in Table VII. By comparing the performance of various approaches on identical datasets, we observe the following three trends:

(1) Generally, incorporating more diverse modalities leads to improved experimental outcomes. For instance, fusing audio, video, and text data yields superior results compared to two-modality or single-modality data. Different modalities increase feature diversity, enabling more effective feature selection for depression prediction.

(2) SVR outperforms other traditional machine learning-

TABLE VII
EXPERIMENTAL RESULTS BASED ON MIXED DATA. WE SUMMARIZE REPRESENTATIVE MACHINE LEARNING METHODS AND REPORT THE EVALUATION METRICS: F1-SCORE, RMSE, AND MAE.

Paper	Dataset	Methods	Data	F1-Score	RMSE	MAE
Joshi et al. [309]	30D+30C	SVM	AV	91.70(Acc)	-	-
Meng et al. [226]	AVEC2013	Partial least square	AV	-	8.72	10.96
Kachele et al. [313]	AVEC2013	SVR	AV	-	8.97	10.82
Niu et al. [314]	AVEC2013	CNN+LSTM	AV	-	8.16	6.14
Cummins et al. [233]	AVEC2013	SVR	AV	-	-	10.44
Pampouchidou et al. [224]	AVEC2013	Nearest Neighbor	AV	73.00	-	-
AVEC2014 [88]	AVEC2014	SVR	AV	-	7.89	9.89
Humberto et al. [315]	AVEC2014	Meta-model	AV	-	6.79	8.31
Jan et al. [244]	AVEC2014	DCNN	AV	-	7.43	6.14
Yang et al. [316]	DAIC-WOZ	DCNN+CNN	AV	-	6.35	5.39
Chao et al. [241]	AVEC2014	CNN+LSTM	AV	-	9.98	7.91
Niu et al. [314]	AVEC2014	CNN+LSTM	AV	-	7.03	5.21
AVEC2016/2017 [25], [89]	DAIC-WOZ	RF	AV	-	7.05	5.66
Yang et al. [317]	DAIC-WOZ	DCNN+DNN	AV	60.00	6.34	5.39
AVEC2019 [90]	E-DAIC	CNN	AV	-	6.37	-
Bilalpur et al. [310]	75D+77C	SVM	AV	78.20(Acc)	-	-
Alghowinem et al. [318]	BlackDog+ Pitt+AVEC2013	SVM	AV	74.60(Acc)	-	-
Morales et al. [319]	AVEC2014	SVM	AT	-	9.21	7.56
Fraser et al. [320]	196D+128C	SVM	AT	79.90(Acc)	-	-
Alhanai et al. [321]	DAIC-WOZ	LSTM	AT	77.00	6.37	5.10
Lau et al. [322]	DAIC-WOZ	Bi-directional LSTM	AT	92.00	3.95	3.00
Lam et al. [323]	DAIC-WOZ	CNN+Transformer	AT	87.00	-	-
Niu et al. [312]	DAIC-WOZ	GAN	AT	92.00	3.80	2.94
Toto et al. [324]	DAIC-WOZ	LSTM+Transformer	AT	92.00	-	-
Flores et al. [325]	DAIC-WOZ	LSTM+Transformer	AT	93.00	-	-
Williamson et al. [326]	DAIC-WOZ	Gaussian Staircase	AVT	81.00	5.31	4.18
Gong et al. [311]	DAIZ-WOZ	Stochastic gradient descent	AVT	60.00	4.99	3.96
Yang et al. [84]	DAIC-WOZ	DCNN+DNN	AVT	-	5.40	4.36
Yang et al. [327]	DAIC-WOZ	DCNN+DNN	AVT	-	5.97	5.16
Yin et al. [328]	E-DAIC	LSTM	AVT	-	5.50	-
Rodrigues et al. [257]	E-DAIC	LSTM+Transformer +CNN+Gated CNN	AVT	-	6.11	-
Morales et al. [329]	DAIC-WOZ	SVM	AT	49.00	-	-
			AV	49.00	-	-
			ATV	49.00	-	-
Rohanian et al. [330]	DAIC-WOZ	LSTM	AT	80.00-	5.14	3.66
			AVT	81.00	4.99	3.61
Qureshi et.al. [331]	DAIC-WOZ	LSTM	AV	-	5.25	3.89
			VT	-	5.11	3.65
			AT	-	4.64	3.65
			AVT	-	4.14	3.07
Ray et al. [250]	E-DAIC+ DAIC-WOZ	LSTM	VT	-	4.64	-
			AT	-	4.37	-
			AVT	-	4.28	-

[1] People with depression (D) and healthy controls (C).
[2] Accuracy (Acc).

based methods on the AVEC2013/2014 dataset [35], [88], with the best results of MAE 10.44 and 9.89 [88], [233], respectively. However, its true improvement in performance lies in aspects such as feature extraction and label processing. First, Local Gabor Binary Pattern Three Orthogonal Planes features have demonstrated better performance than other descriptors in human behavior analysis. Second, averaging dimensional affect labels across multiple subjective ratings helps mitigate subjectivity in affective behavior interpretation and reduces rater errors caused by cognitive workload effects such as fatigue.

(3) On the DAIC-WOZ dataset [99], stochastic gradient descent regression outperforms other traditional machine learning-based methods with a MAE of 3.96. Gong et al. [311] proposed a topic modeling-based multimodal feature vector construction scheme for context-aware analysis and conducted a grid search on RF regression, stochastic gradient descent regression, and SVR models. The authors observe that as feature numbers increase, stochastic gradient descent and SVR models consistently improve their performance, while the random forest model plateaus earlier. The results indicate that the proposed approach significantly outperforms both the context-unaware method and the challenge baseline across all metrics , uncovering various temporal features with underlying relationships. Additionally, the attention-based approach that combines audio and text modalities achieved the best result, with a MAE of 2.94 [312]. This improvement can be attributed to the attention mechanism's ability to enhance the model's understanding of the relationships between different modalities of data.

VIII. FUTURE DIRECTION

Following a comprehensive review of current approaches for depression prediction based on different data modalities, this study presents future research directions from six perspectives: task, dataset, model, application, inter-modal comparisons, and ethical issue.

A. Task

To fully and comprehensively investigate depression, several new character variables must be addressed and resolved.

1) Subtypes of depression: The existing research often distinguishes between depressed and nondepressed individuals and predicts depression severity, but it usually overlooks the distinct subtypes of depression. For instance, bipolar disorder and major depressive disorder, which is the most common subtype, differ in their origins and behaviors [332]. Depression also closely intertwines with other mental illnesses like Alzheimer's disease and post-traumatic stress disorder due to shared emotional and pathological connections, frequently coexisting with depression [333], [334]. However, each illness has its own unique causes, pathologies, and diagnostic criteria, despite symptom similarities. In addition, the depression scale can offer detailed insights through analysis. Each question represents a specific symptom, and analyzing individual responses provides a comprehensive understanding of the subject's psychological state [282]. Yet, research comparing these subtypes remains scarce. Developing tailored models for different depression subtypes could improve depression prediction and provide more accurate clinical diagnosis and treatment guidance.

2) Longitudinal study: Current research on depression predominantly relies on cross-sectional studies [335], [336], with limited longitudinal research that follows patients from the onset to remission of the illness. Moreover, most existing longitudinal studies primarily use basic statistical analyses [337]–[340], overlooking the potential of machine learning techniques. These methods offer significant advantages: first, they can automatically identify critical features, including temporal dynamics, shedding light on the development and influencing factors of depression. Second, machine learning methods can tailor analyses to individual longitudinal data, highlighting differences among subgroups. Integrating machine learning into longitudinal studies could lead to more precise and effective predictions for depression. Understanding the entire course of the illness is essential to uncover the neuropathological mechanisms of depression. Future research should therefore prioritize longitudinal studies using machine learning methods to deepen our understanding of the pathology behind depression's onset and progression.

3) Treatment: In addition to specialized medications and psychotherapy for depression, numerous other approaches exist, including physical interventions (transcranial magnetic stimulation and electroconvulsive shock), exercise and music therapy. These modalities are designed to promote relaxation and mood regulation in patients, with the ultimate goal of ameliorating depressive symptoms. Future research could evaluate the efficacy of various treatment modalities in facilitating recovery from depression, with the aim of identifying the most effective treatment options.

B. Datasets

Although depression prediction has progressed with existing datasets, there is still space for advancement in terms of diversity, clinicality, and publicity.

1) Diversity: Past depression prediction studies often used small sample sizes, limiting the accuracy and generalizability of their models. To build a larger and more comprehensive depression database, multiple factors must be considered [341]: **(1) Background information of the subjects.** Depression's manifestation varies with age, gender, cultural background, family dynamics, and disease state [45], [342]–[344]. Investigating extensive datasets encompassing diverse patient demographics can enhance classification performance and provide a thorough analysis of how these factors influence depression. **(2) Collection scene.** It is important to establish an environment of trust and safety with patients when collecting data. In future studies, patients can be helped to feel more comfortable and relaxed by constructing comfortable data collection scenarios and incorporating human interactions, in order to allow them to show their depressive symptoms more naturally. **(3) Data for more modalities.** Depression's complexity necessitates the collection and analysis of various psychological and physiological indicators for an objective assessment. Future research should aim to gather multimodal data, including brain imaging, EEG, and audio/video recordings, for joint analysis. **(4) Quality of data.** The collection and integration of multimodal data is a complex process that requires facing various technical and methodological challenges. Future research must address issues like missing data, noise, and inconsistencies to improve data quality and model credibility.

2) Clinicality: To conduct machine learning-based depression prediction research, a comprehensive understanding of the medical mechanism of depression is essential. **(1) Clinical information.** Clinical information about the patient can be collected, such as medical history, symptoms, physiologic indicators, medications, and psychological assessments. These information can be used to construct analytic models that help predict depression. **(2) Involvement of professionals.** Depression is a complex disorder that requires the combined efforts of physicians, psychologists, nurses and other healthcare professionals for comprehensive treatment. Through enhanced multidisciplinary cooperation and coordination, the expertise and experience of each field can be better utilized to provide more comprehensive and effective support for the prediction and treatment of depression.

3) Publicity: Due to the sensitivity of depression data and ethical issues, most institutions cannot access adequate samples. There are few publicly available databases on depression, which leaves a large gap in scientific research. Establishing an open-access depression database with international contributors could potentially increase sample sizes for future studies. Collaborative efforts from medical experts and researchers worldwide could generate much larger datasets, ultimately improving the reliability of models for clinical trial use.

C. Model

Several factors impact model performance, including unbalanced samples, model interpretability, applicability, and multimodal fusion. In this study, we provide insights into potential solutions for these issues.

1) Unbalanced samples: Addressing sample imbalance, a prevalent issue in depression prediction studies, especially in data-driven deep learning. To address this challenge, researchers can consider multiple strategies: **(1) Data augmentation.** Generative models [190], particularly effective in balancing training data, create new samples by interpolating within semantic or feature spaces. In semantic space, interpolation occurs through higher-level dimensions like emotions or intentions, ensuring synthetic samples retain semantic consistency with real data. In feature space, interpolations focus on aspects such as pixel values or frequency distributions, ensuring generated samples are perceptually akin to the original in visual or auditory attributes. This approach yields high-quality synthetic data that maintains the original semantics and features, offering a robust solution to data imbalance. **(2) Knowledge migration.** Knowledge migration shows its potential when dealing with unbalanced data. When faced with data scarcity, researchers may consider introducing pre-trained models that are relevant to the target task. For instance, in depression prediction, weights and features from abundant emotion analysis data and models can be transferred, balancing the dataset with knowledge from another domain and enhancing model accuracy and robustness. **(3) Sample re-weighting.** Assigning greater weights to a small number of categories can reinforce the model's learning of these key samples, thereby improving the accuracy of depression prediction.

2) Interpretability: Deep learning-based approaches for depression prediction typically outperform traditional machine learning methods, but may lack interpretability. To enhance the interpretability of the model, researchers can adopt the following strategy: **(1) Interpretable Models.** First, fusing artificial feature extraction with neural networks. This involves merging artificial feature extraction from traditional machine learning with deep learning models, which not only leverages the deep learning models' capability to extract intricate features from data but also assimilates human expert domain knowledge. This dual approach enhances both the performance and interpretability of the model. Second, implementing attention mechanisms. Attention mechanisms allow the model to concentrate on essential parts of the input data, aiding in pinpointing the model's focus areas during predictions, such as paying more attention to text or features associated with mood during depression prediction. **(2) Model Agnostic Methods.** Model agnostic methods do not take into account the structure of the model. They obtain explanations by perturbing and mutating the input data and obtaining sensitivity of the performance of these mutations with respect to the original data performance. They help reveal the model's prediction logic based on different features and variations, which is crucial for diagnosing complex disorders like depression. **(3) Example-based Explanations.** Example-based explanation methods select particular instances of the dataset to explain the behaviour of machine learning models or the underlying data distribution. In depression prediction, feature visualization offers a graphical analysis of individual sample characteristics, elucidating the model's decision-making rationale. Through visualization, medical professionals and patients are more inclined to accept the predictive outcomes, bolstering the model's transparency and trustworthiness.

3) Multimodal fusion: Data from a single modality may not offer a comprehensive insight into the features of depression. Conversely, multimodal data can enhance the scope of input data by leveraging complementary information, thereby bolstering the performance of depression prediction. Future research can focus on the following aspects: **(1) Exploring Multimodal Data Interrelationships.** Investigating the interrelationships between multimodal data is a central challenge in current and future research domains. Notably, it is pivotal in revealing the underlying connections between different data modalities and enhancing our recognition and understanding of mental illnesses such as depression. For instance, by synthesizing data from various modalities, such as physiological signals, facial expressions, textual descriptions, and vocal characteristics of individuals with depression, researchers can delve into their interrelations and potential correlations. **(2) Developing Multimodal Fusion Techniques.** Depression prediction is a complex task that requires integrating data from various modalities. Therefore, it is crucial to develop advanced multimodal fusion strategies. For example, early fusion integrates different modalities at the initial stage, allowing information to be synthesized more directly. Conversely, late fusion synthesizes insights by combining the results of each modality's independent analyses during the final stages. Additionally, model-based fusion offers an impressive way to utilize specialized models to identify and assimilate interactions between data sources. These techniques do more than overlay different data modalities. They aim to incorporate a coherent whole capable of depicting depression in all dimensions.

4) Algorithm Scalability: To enhance the relevance of our survey to real-world applications in depression prediction, it is essential to consider the scalability and adaptability of machine learning and deep learning algorithms in practical settings. Evaluating these methods based on computational complexity, resource requirements, and their potential for distributed implementation can provide a clearer understanding of how these models perform in applied environments. In the following, we discuss three key aspects: **(1) Complexity and Resource Demand of Machine Learning.** The complexity and resource requirements of machine learning algorithms are critical for their feasibility in real-world applications like depression prediction. Simpler models require minimal processing power and memory, making them suitable for resource-constrained environments. However, as models grow more complex to handle diverse and larger datasets, their computational demands increase, which can pose challenges during development and deployment. Furthermore, when machine learning algorithms are applied to multimodal data or used in dynamic and interactive applications, resource requirements tend to increase due to the need for complex data preprocessing and integration. Efficient model design and optimization are key to keeping these algorithms practical for real-world use. **(2) Potential for Distributed and Scalable Implementation.** In real-world applications, distributed computing frameworks and cloud-based solutions offer promising ways to enhance the scalability of deep learning approaches. Leveraging distributed training can accelerate the model development process and make it feasible to handle large datasets. Similarly, cloud computing platforms allow for on-demand scaling of computational resources, enabling deep learning models to be deployed effectively even in resource-intensive tasks. This approach is increasingly seen in healthcare settings, where real-time and large-scale applications like depression predicting benefit from such scalable infrastructure. **(3) Optimizing Computational Efficiency for Practical Deployment.** Improving model efficiency is key to enabling deep learning algorithms to meet real-world requirements. By exploring strategies like model pruning, quantization, or designing lightweight architectures, resource consumption can be significantly reduced. These optimization techniques can enhance the applicability of deep learning models in settings with limited resources, expanding the reach of tasks such as depression prediction to broader healthcare contexts and diverse operational environments.

D. Application

We propose several future research directions for the innovative application of depression prediction methods:

1) Clinical application: Depression often goes unrecognized in clinical diagnosis, with many mistakenly believing that patients are free of disease, leading to delayed diagnosis and treatment. To address it, data-driven behavioral depression prediction methods can be employed to aid in patient diagnosis. Furthermore, depression prediction methods can be utilized to assess changes in patients' physical condition over time and to evaluate the effectiveness of treatments and other factors in promoting recovery. Such methods can provide valuable support in identifying effective treatments and ensuring that patients receive appropriate care.

Existing studies analyzing depression based on data from various modalities have achieved good results, but they are rarely applied to clinical analysis, mainly due to the following reasons: **(1) Limited data on clinically depressed patients.** Currently, there are relatively few data on clinically depressed patients, which makes it difficult to get enough clinical data to support the study of analyzing depression based on different modality data. **(2) Data heterogeneity.** Data from different sources may display inconsistencies. It poses challenges to the training and generalization of machine learning models. Particularly in clinical settings, this inconsistency may be further amplified due to individual differences in patients, increasing the risk of misdiagnosis. **(3) Reproducibility and reliability.** Notably, some studies use private data, limiting comparisons with other research. Additionally, a lack of transparency in data processing, feature selection, and model parameters casts doubts on their reproducibility and reliability. Research on depression analysis based on different modalities is not yet mature and and lacks a unified framework. While deep learning methods have shown promising results, their clinical applicability is limited due to their 'black box' nature and lack of interpretability. These results can also be influenced by factors like data bias and model complexity, necessitating further refinement for clinical use.

2) Daily self-assessment: Depression is a prevalent mental illness often characterized by abnormal emotions. It has caused a significant increase in depression across all age groups. Daily assessment using easily collected behavioral data provides real-time insight into one's condition and allows prompt medical attention when abnormalities are detected. Additionally, complementary depression prediction methods can be utilized for mass screening in colleges, communities, and other crowded areas, enabling the early diagnosis of individuals who may be suffering from depression.

E. Ethical issue

In the field of depression prediction, ethical issues are of paramount importance. When conducting research, it is imperative to uphold ethical principles and ensure the privacy and rights of subjects are fully respected. Further, ethical considerations vary when dealing with different data modalities.

1) Biological signal data: When processing biological signal data, such as electrophysiological signals and brain imaging, it is necessary to strictly adhere to medical ethics and privacy regulations to ensure the secure storage and use of data. Firstly, it is crucial to ensure that participants are fully informed about the purpose, use, and potential risks associated with providing biometric signal data. Secondly, for sensitive data that may reveal personal health information, higher standards of privacy protection measures should be employed, such as enhanced data anonymization and secure encryption technologies. Finally, considering the complex medical ethical issues associated with biometric signal data, such as predictions about future health states, close collaboration with an ethics committee should be established from the design phase of the research to ensure that all procedures comply with ethical norms.

2) Behavioral data: For behavioral data, particularly those containing potential personal identifiers such as facial images and voice recordings, stringent privacy protection measures must be implemented. All collected data undergo a process of de-identification before use, ensuring that individual identities cannot be directly or indirectly discerned.

3) Text: When utilizing publicly available textual data from sources such as social media, it is imperative to consider individual privacy rights. Firstly, it is crucial to pay attention to potentially sensitive information in the text, such as details of personal identity and health status, to ensure that such information is not disclosed during the analysis and reporting process. Secondly, a clear distinction must be made between data that is publicly accessible and data that requires special authorization for use, while respecting the copyright and usage regulations of the original data sources.

In summary, when studying depression prediction, it is necessary to strictly adhere to ethical principles and take appropriate protective measures for different types of data to ensure the fairness and validity of research.

IX. Conclusion

In a study commissioned by the World Health Organization, it was determined that the prevalence of depression and anxiety escalated by more than 25% during the pandemic. This finding underscores the necessity for scholarly investigation and clinical attention to devise an automated and objective assessment system. This paper offers a comprehensive and thorough examination of the literature with a focus on machine learning applications in various data modalities. The paper encapsulates the general research process and typical methodologies employed in machine learning for the prediction of depression. Upon a meticulous evaluation of extant works, we propose a range of future research avenues. This contribution is aimed at assisting psychiatric researchers in developing more dependable and sophisticated systems.

References

[1] Q. Wang, "The social determinants of depressive disorders in china," *Lancet psychiatry*, 2021.

[2] D. Bhugra et al., "Globalisation and mental disorders: overview with relation to depression," *The British Journal of Psychiatry*, vol. 184, no. 1, pp. 10–20, 2004.

[3] T. S. Rao et al., "Understanding nutrition, depression and mental illnesses," *Indian journal of psychiatry*, vol. 50, no. 2, p. 77, 2008.

[4] W. H. Organization. (2021) Depression. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[5] R. Peveler et al., "Depression in medical patients," *Bmj*, vol. 325, no. 7356, pp. 149–152, 2002.

[6] K. Demyttenaere et al., "Comorbid painful physical symptoms and depression: prevalence, work loss, and help seeking," *Journal of affective disorders*, vol. 92, no. 2-3, pp. 185–193, 2006.

[7] A. Singh et al., "Loneliness, depression and sociability in old age," *Industrial psychiatry journal*, vol. 18, no. 1, p. 51, 2009.

[8] D. Eisenberg et al., "Prevalence and correlates of depression, anxiety, and suicidality among university students," *American journal of orthopsychiatry*, vol. 77, no. 4, pp. 534–542, 2007.

[9] I. O. Bergfeld et al., "Treatment-resistant depression and suicidality," *Journal of affective disorders*, vol. 235, pp. 362–367, 2018.

[10] L. Hemming et al., "Alexithymia and its associations with depression, suicidality, and aggression: an overview of the literature," *Frontiers in psychiatry*, vol. 10, p. 203, 2019.

[11] J. X. Wiebenga et al., "Associations of three major physiological stress systems with suicidal ideation and suicide attempts in patients with a depressive and/or anxiety disorder," *Brain, behavior, and immunity*, vol. 102, pp. 195–205, 2022.

[12] D. Faust et al., "The expert witness in psychology and psychiatry," *Science*, vol. 241, no. 4861, pp. 31–35, 1988.

[13] P. Cassano et al., "Depression and public health: an overview," *Journal of psychosomatic research*, vol. 53, no. 4, pp. 849–857, 2002.

[14] D. C. Mohr et al., "Barriers to psychotherapy among depressed and nondepressed primary care patients," *Annals of Behavioral Medicine*, vol. 32, no. 3, pp. 254–258, 2006.

[15] R. L. Kravitz et al., "Relational barriers to depression help-seeking in primary care," *Patient education and counseling*, vol. 82, no. 2, pp. 207–213, 2011.

[16] http://www.cbsr.ia.ac.cn/users/ynyu/dataset/.

[17] N. Cheng et al., "Addressing the nationwide shortage of child and adolescent psychiatrists: determining factors that influence the decision for psychiatry residents to pursue child and adolescent psychiatry training," *Academic psychiatry*, vol. 46, no. 1, pp. 18–24, 2022.

[18] T. Butryn et al., "The shortage of psychiatrists and other mental health providers: causes, current state, and potential solutions," *International Journal of Academic Medicine*, vol. 3, no. 1, p. 5, 2017.

[19] S. Abuse *et al.*, "Key substance use and mental health indicators in the united states: results from the 2019 national survey on drug use and health," 2020.

[20] B. A. Pescosolido et al., "The "backbone of stigma: identifying the global core of public prejudice associated with mental illness," *American journal of public health*, vol. 103, no. 5, pp. 853–860, 2013.

[21] S. Wang, X. Zhu, W. Ding, and A. A. Yengejeh, "Cyberbullying and cyberviolence detection: A triangular user-activity-content view," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 8, pp. 1384–1405, 2022.

[22] U. R. Acharya et al., "Automated eeg-based screening of depression using deep convolutional neural network," *Computer methods and programs in biomedicine*, vol. 161, pp. 103–113, 2018.

[23] R. Wang et al., "Depression analysis and recognition based on functional near-infrared spectroscopy," *IEEE journal of biomedical and health informatics*, vol. 25, no. 12, pp. 4289–4299, 2021.

[24] W. Melo et al., "Combining global and local convolutional 3d networks for detecting depression from facial expressions," in *2019 14th ieee international conference on automatic face & gesture recognition (fg 2019)*. IEEE, 2019, pp. 1–8.

[25] F. Ringeval et al., "Avec 2017: Real-life depression, and affect recognition workshop and challenge," in *Proceedings of the 7th annual workshop on AVEC*, 2017, pp. 3–9.

[26] H. Lu et al., "Postgraduate student depression assessment by multimedia gait analysis," *IEEE MultiMedia*, 2022.

[27] G. Rao et al., "Mgl-cnn: a hierarchical posts representations model for identifying depressed individuals in online forums," *IEEE Access*, vol. 8, pp. 32 395–32 403, 2020.

[28] L. Jin, S. Li, J. Yu, and J. He, "Robot manipulator control using neural networks: A survey," *Neurocomputing*, vol. 285, pp. 23–34, 2018.

[29] Y. Zhang, S. Li, and X. Zhou, "Recurrent-neural-network-based velocity-level redundancy resolution for manipulators subject to a joint acceleration limit," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3573–3582, 2018.

[30] S. Li, Z. Shao, and Y. Guan, "A dynamic neural network approach for efficient control of manipulators," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 5, pp. 932–941, 2017.

[31] J. Li, Y. Zhang, S. Li, and M. Mao, "New discretization-formula-based zeroing dynamics for real-time tracking control of serial and parallel manipulators," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3416–3425, 2017.

[32] J. Luo et al., "Big data application in biomedical research and health care: a literature review," *Biomedical informatics insights*, vol. 8, pp. BII–S31 559, 2016.

[33] S.-C. Liao et al., "Major depression detection from eeg signals using kernel eigen-filter-bank common spatial patterns," *Sensors*, vol. 17, no. 6, p. 1385, 2017.

[34] M. Wei et al., "Identifying major depressive disorder using hurst exponent of resting-state brain networks," *Psychiatry Research: Neuroimaging*, vol. 214, no. 3, pp. 306–312, 2013.

[35] M. Valstar et al., "Avec 2013: the continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, 2013, pp. 3–10.

[36] S. Tsugawa et al., "Recognizing depression from twitter activity," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 3187–3196.

[37] G. Xu, A. S. Khan, A. J. Moshayedi, X. Zhang, and Y. Shuxin, "The object detection, perspective and obstacles in robotic: a review," *EAI Endorsed Transactions on AI and Robotics*, vol. 1, no. 1, 2022.

[38] A. J. Moshayedi, A. S. Khan, S. Yang, and S. M. Zanjani, "Personal image classifier based handy pipe defect recognizer (hpd): Design and test," in *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP)*. IEEE, 2022, pp. 1721–1728.

[39] G. Sharma et al., "Dephnn: a novel hybrid neural network for electroencephalogram (eeg)-based screening of depression," *Biomedical signal processing and control*, vol. 66, p. 102393, 2021.

[40] M. Mousavian et al., "Depression detection using atlas from fmri images," in *19th IEEE ICMLA*, 2020, pp. 1348–1353.

[41] P. Zhang et al., "Depa: Self-supervised audio embedding for depression detection," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 135–143.

[42] J. Yang et al., "Data augmentation for depression detection using skeleton-based gait information," *Medical & Biological Engineering & Computing*, vol. 60, no. 9, pp. 2665–2679, 2022.

[43] U. Lee, G. Jung, E.-Y. Ma, J. San Kim, H. Kim, J. Alikhanov, Y. Noh, and H. Kim, "Toward data-driven digital therapeutics analytics: Literature review and research directions," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 1, pp. 42–66, 2023.

[44] K. M. Hasib, M. R. Islam, S. Sakib, M. A. Akbar, I. Razzak, and M. S. Alam, "Depression detection from social networks data based on machine learning and deep learning techniques: An interrogative survey," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 4, pp. 1568–1586, 2023.

[45] L. He et al., "Deep learning for depression recognition with audiovisual cues: A review," *Information Fusion*, vol. 80, pp. 56–86, 2022.

[46] S. Chancellor et al., "Methods in predictive techniques for mental health status on social media: a critical review," *NPJ digital medicine*, vol. 3, no. 1, p. 43, 2020.

[47] A. Safayari et al., "Depression diagnosis by deep learning using eeg signals: A systematic review," *Medicine in Novel Technology and Devices*, vol. 12, p. 100102, 2021.

[48] A. Ashraf et al., "On the review of image and video-based depression detection using machine learning," *IJEECS*, vol. 19, no. 3, pp. 1677–1684, 2020.

[49] S. Aleem et al., "Machine learning algorithms for depression: diagnosis, insights, and research directions," *Electronics*, vol. 11, no. 7, p. 1111, 2022.

[50] P. Wu et al., "Automatic depression recognition by intelligent speech signal processing: A systematic survey," *CAAI Transactions on Intelligence Technology*, vol. 8, no. 3, pp. 701–711, 2023.

[51] A. Sarkar et al., "A deep learning-based comparative study to track mental depression from eeg data," *Neuroscience Informatics*, vol. 2, no. 4, p. 100039, 2022.

[52] R. Salas-Zárate et al., "Detecting depression signs on social media: a systematic literature review," in *Healthcare*, vol. 10, no. 2. MDPI, 2022, p. 291.

[53] S. J. Pinto and M. Parente, "Comprehensive review of depression detection techniques based on machine learning approach," *Soft Computing*, pp. 1–25, 2024.

[54] S. Bhadra et al., "An insight into diagnosis of depression using machine learning techniques: a systematic review," *Current medical research and opinion*, vol. 38, no. 5, pp. 749–771, 2022.

[55] A. Widmann et al., "Digital filter design for electrophysiological data–a practical approach," *Journal of neuroscience methods*, vol. 250, pp. 34–46, 2015.

[56] D. M. Schnyer et al., "Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor mri measures in individuals with major depressive disorder," *Psychiatry Research: Neuroimaging*, vol. 264, pp. 1–9, 2017.

[57] J. R. Sato et al., "Machine learning algorithm accurately detects fmri signature of vulnerability to major depression," *Psychiatry Research: Neuroimaging*, vol. 233, no. 2, pp. 289–291, 2015.

[58] R. Ramasubbu et al., "Accuracy of automated classification of major depressive disorder as a function of symptom severity," *NeuroImage: Clinical*, vol. 12, pp. 320–331, 2016.

[59] B. Vai et al., "Predicting differential diagnosis between bipolar and unipolar depression with multiple kernel learning on multimodal structural neuroimaging," *European Neuropsychopharmacology*, vol. 34, pp. 28–38, 2020.

[60] Y. Wei et al., "Functional near-infrared spectroscopy (fnirs) as a tool to assist the diagnosis of major psychiatric disorders in a chinese population," *European archives of psychiatry and clinical neuroscience*, vol. 271, no. 4, pp. 745–757, 2021.

[61] Y. Wang, S. Qiu, D. Li, C. Du, B.-L. Lu, and H. He, "Multi-modal domain adaptation variational autoencoder for eeg-based emotion recognition," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 9, pp. 1612–1626, 2022.

[62] A. Lenartowicz et al., "Brain imaging," 2017.

[63] E. J. Nestler et al., "Neurobiology of depression," *Neuron*, vol. 34, no. 1, pp. 13–25, 2002.

[64] G. Nilsonne et al., "Eeg-based model and antidepressant response," *Nature Biotechnology*, vol. 39, no. 1, pp. 27–27, 2021.

[65] M. Shim et al., "Machine-learning-based classification between post-traumatic stress disorder and major depressive disorder using p300 features," *NeuroImage: Clinical*, vol. 24, p. 102001, 2019.

[66] H. Jiang et al., "Predictability of depression severity based on posterior alpha oscillations," *Clinical Neurophysiology*, vol. 127, no. 4, pp. 2108–2114, 2016.

[67] X. Li et al., "Classification study on eye movement data: Towards a new approach in depression detection," in *IEEE CEC*, 2016, pp. 1227–1232.

[68] A. Y. Kim et al., "Automatic detection of major depressive disorder using electrodermal activity," *Scientific reports*, vol. 8, no. 1, pp. 1–9, 2018.

[69] C. Sobin et al., "Psychomotor symptoms of depression," *American Journal of Psychiatry*, vol. 154, no. 1, pp. 4–17, 1997.

[70] X. Ma et al., "Depaudionet: An efficient deep model for audio based depression classification," in *Proceedings of the 6th international workshop on AVEC*, 2016, pp. 35–42.

[71] X. Zhou et al., "Visually interpretable representation learning for depression recognition from facial images," *IEEE transactions on affective computing*, vol. 11, no. 3, pp. 542–552, 2018.

[72] B. Miao et al., "Automatic mental health identification method based on natural gait pattern," *PsyCh Journal*, vol. 10, no. 3, pp. 453–464, 2021.

[73] Z. Liu, M. Wu, W. Cao, L. Chen, J. Xu, R. Zhang, M. Zhou, and J. Mao, "A facial expression emotion recognition based human-robot interaction system." *IEEE CAA J. Autom. Sinica*, vol. 4, no. 4, pp. 668–676, 2017.

[74] S. Thomée et al., "Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults-a prospective cohort study," *BMC public health*, vol. 11, no. 1, pp. 1–11, 2011.

[75] Z. Huang et al., "Depression detection from short utterances via diverse smartphones in natural environmental conditions." in *INTERSPEECH*, 2018, pp. 3393–3397.

[76] A. Dogrucu et al., "Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data," *Smart Health*, vol. 17, p. 100118, 2020.

[77] M. Trotzek et al., "Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences," *IEEE TKDE*, vol. 32, no. 3, pp. 588–601, 2018.

[78] S. Park et al., "Manifestation of depression and loneliness on social networks: a case study of young adults on facebook," in *ACM CSCW*, 2015, pp. 557–570.

[79] X. Yang et al., "A big data analytics framework for detecting user-level depression from social networks," *IJIM*, vol. 54, p. 102141, 2020.

[80] X. Kang, F. Ren, and Y. Wu, "Exploring latent semantic information for textual emotion recognition in blog articles," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 204–216, 2017.

[81] P. A. Lalousis et al., "Heterogeneity and classification of recent onset psychosis and depression: a multimodal machine learning approach," *Schizophrenia bulletin*, vol. 47, no. 4, pp. 1130–1140, 2021.

[82] Y. Meng et al., "Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression," *IEEE J BIOMED HEALTH*, vol. 25, no. 8, pp. 3121–3129, 2021.

[83] Y. J. Toenders et al., "Predicting depression onset in young people based on clinical, cognitive, environmental, and neurobiological data," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 7, no. 4, pp. 376–384, 2022.

[84] L. Yang et al., "Hybrid depression classification and estimation from audio video and text information," in *Proceedings of the 7th annual workshop on AVEC*, 2017, pp. 45–51.

[85] U. R. Acharya et al., "A novel depression diagnosis index using nonlinear features in eeg signals," *European neurology*, vol. 74, no. 1-2, pp. 79–83, 2015.

[86] H. Cai et al., "A multi-modal open dataset for mental-disorder analysis," *Scientific Data*, vol. 9, no. 1, p. 178, 2022.

[87] S. Hong et al., "Automatic depression discrimination on fnirs by using fastica/wpd and svm," in *Proceedings of the 2015 Chinese Intelligent Automation Conference: Intelligent Information Processing*. Springer, 2015, pp. 257–265.

[88] M. Valstar et al., "Avec 2014: 3d dimensional affect and depression recognition challenge," in *Proceedings of the 4th international workshop on AVEC*, 2014, pp. 3–10.

[89] ——, "Avec 2016: Depression, mood, and emotion recognition workshop and challenge," in *Proceedings of the 6th international workshop on AVEC*, 2016, pp. 3–10.

[90] F. Ringeval et al., "Avec 2019 workshop and challenge: state-of-mind, detecting depression with ai, and cross-cultural affect recognition," in *Proceedings of the 9th International on AVEC and Workshop*, 2019, pp. 3–12.

[91] Y. Shen et al., "Automatic depression detection: An emotional audio-textual corpus and a gru/bilstm-based model," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 6247–6251.

[92] A. Yates et al., "Depression and self-harm risk assessment in online forums," *arXiv preprint arXiv:1709.01848*, 2017.

[93] M. Stankevich et al., "Feature engineering for depression detection in social media," in *ICPRAM*, 2018.

[94] G. Coppersmith et al., "Clpsych 2015 shared task: Depression and ptsd on twitter," in *Proceedings of the 2nd workshop on CLPsych: from linguistic signal to clinical reality*, 2015, pp. 31–39.

[95] Z. Jamil, "Monitoring tweets for depression to detect at-risk users," Ph.D. dissertation, Université d'Ottawa/University of Ottawa, 2017.

[96] Y. Yuan et al., "Depression identification from gait spectrum features based on hilbert-huang transform," in *HCC*. Springer, 2018, pp. 503–515.

[97] F. Ringeval et al., "Avec 2015: The 5th international audio/visual emotion challenge and workshop," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 1335–1336.

[98] ——, "Avec 2018 workshop and challenge: Bipolar disorder and cross-cultural affect recognition," in *Proceedings of the 2018 on audio/visual emotion challenge and workshop*, 2018, pp. 3–13.

[99] J. Gratch et al., "The distress analysis interview corpus of human and computer interviews," UNIVERSITY OF SOUTHERN CALIFORNIA LOS ANGELES, Tech. Rep., 2014.

[100] C.-h. Wuu et al., "Multiface: A dataset for neural face rendering," *arXiv preprint arXiv:2207.11243*, 2022.

[101] M. Tlachac et al., "Emu: Early mental health uncovering framework and dataset," in *2021 20th IEEE ICMLA*, 2021, pp. 1311–1318.

[102] S. Alghowinem et al., "Multimodal depression detection: fusion analysis of paralinguistic, head pose and eye gaze behaviors," *IEEE TAC*, vol. 9, no. 4, pp. 478–490, 2016.

[103] Y. Yang et al., "Detecting depression severity from vocal prosody," *IEEE transactions on affective computing*, vol. 4, no. 2, pp. 142–150, 2012.

[104] D. Highland et al., "A review of detection techniques for depression and bipolar disorder," *Smart Health*, vol. 24, p. 100282, 2022.

[105] Z.-x. Yang et al., "An effective sparsity evaluation criterion for power-line interference suppression of eeg signal," *Frontiers in Neuroscience*, vol. 16, p. 984471, 2022.

[106] Y. Sun et al., "Capacitive biopotential measurement for electrophysiological signal acquisition: A review," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 2832–2853, 2016.

[107] L. Xu et al., "Use of power-line interference for adaptive motion artifact removal in biopotential measurements," *Physiological measurement*, vol. 37, no. 1, p. 25, 2015.

[108] S. B. Kotsiantis et al., "Data preprocessing for supervised leaning," *International journal of computer science*, vol. 1, no. 2, pp. 111–117, 2006.

[109] S. García et al., *Data preprocessing in data mining*. Springer, 2015, vol. 72.

[110] X. Gong et al., "Joint prediction of multiple vulnerability characteristics through multi-task learning," in *2019 24th ICECCS*, 2019, pp. 31–40.

[111] M. M. Meskhi et al., "Learning abstract task representations," in *AAAI Workshop on Meta-Learning and MetaDL Challenge*. PMLR, 2021, pp. 127–137.

[112] I. Tavchioski et al., "Early detection of depression with linear models using hand-crafted and contextual features," *Working Notes of CLEF*, pp. 5–8, 2022.

[113] J. R. Williamson et al., "Vocal biomarkers of depression based on motor incoordination," in *Proceedings of the 3rd ACM international workshop on AVEC*, 2013, pp. 41–48.

[114] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[115] A. J. Smola et al., "A tutorial on support vector regression," *Statistics and computing*, vol. 14, pp. 199–222, 2004.

[116] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[117] S. Menard, *Applied logistic regression analysis*. Sage, 2002, no. 106.

[118] Y. Kang et al., "Deep transformation learning for depression diagnosis from facial images," in *Chinese conference on biometric recognition*. Springer, 2017, pp. 13–22.

[119] A. Salekin et al., "A weakly supervised learning framework for detecting social anxiety and depression," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 2, no. 2, pp. 1–26, 2018.

[120] T. Ran et al., "Identifying depressive disorder with sleep electroencephalogram data: A study based on deep learning," *Journal of Sichuan University (Medical Science Edition)*, vol. 54, no. 2, 2023.

[121] Y. LeCun et al., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[122] J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.

[123] S. Hochreiter et al., "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[124] A. Vaswani et al., "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[125] Z. Li, S. Li, O. O. Bamasag, A. Alhothali, and X. Luo, "Diversified regularization enhanced training for effective manipulator calibration," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 11, pp. 8778–8790, 2022.

[126] Z. Li, S. Li, A. Francis, and X. Luo, "A novel calibration system for robot arm via an open dataset and a learning perspective," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 12, pp. 5169–5173, 2022.

[127] Z. Li, S. Li, and X. Luo, "An overview of calibration technology of industrial robots," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 1, pp. 23–36, 2021.

[128] E. Aghajari and A. A. AbdulRahim, "Prediction of short circuit current of wind turbines based on artificial neural network model," *EAI Endorsed Transactions on AI and Robotics*, vol. 3, 2024.

[129] M. Davari, A. Harooni, A. Nasr, K. Savoji, and M. Soleimani, "Improving recognition accuracy for facial expressions using scattering wavelet," *EAI Endorsed Transactions on AI and Robotics*, vol. 3, 2024.

[130] A. Gupta, "Improved hybrid preprocessing technique for effective segmentation of wheat canopies in chlorophyll fluorescence images," *EAI Endorsed Transactions on AI and Robotics*, vol. 3, 2024.

[131] X. Luo, W. Qin, A. Dong, K. Sedraoui, and M. Zhou, "Efficient and high-quality recommendations via momentum-incorporated parallel stochastic gradient descent-based learning," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 2, pp. 402–411, 2020.

[132] X. Wen and M. Zhou, "Evolution and role of optimizers in training deep learning models," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 10, pp. 2039–2042, 2024.

[133] W. He, M. Liu, Y. Tang, Q. Liu, and Y. Wang, "Differentiable automatic data augmentation by proximal update for medical image segmentation," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 7, pp. 1315–1318, 2022.

[134] X. Jiang and S. Li, "Bas: Beetle antennae search algorithm for optimization problems. arxiv preprint abs/1710.10724," *URL http://arxiv.org/abs/1710.10724*, vol. 1710, 2017.

[135] Y. Zhang, S. Li, and B. Xu, "Convergence analysis of beetle antennae search algorithm and its applications," *Soft Computing*, vol. 25, no. 16, pp. 10 595–10 608, 2021.

[136] Q. Wu, H. Lin, Y. Jin, Z. Chen, S. Li, and D. Chen, "A new fallback beetle antennae search algorithm for path planning of mobile robots with collision-free capability," *Soft Computing*, vol. 24, pp. 2369–2380, 2020.

[137] A. H. Khan, X. Cao, S. Li, V. N. Katsikis, and L. Liao, "Bas-adam: An adam based approach to improve the performance of beetle antennae search optimizer," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 2, pp. 461–471, 2020.

[138] W. H. Organization. (2022) Depression. [Online]. Available: https://www.who.int/health-topics/depression#tab=tab_1

[139] J. R. Hughes et al., "Conventional and quantitative electroencephalography in psychiatry," *The Journal of neuropsychiatry and clinical neurosciences*, vol. 11, no. 2, pp. 190–208, 1999.

[140] P. L. Faris et al., "Evidence for a vagal pathophysiology for bulimia nervosa and the accompanying depressive symptoms," *Journal of affective disorders*, vol. 92, no. 1, pp. 79–90, 2006.

[141] M. R. Islam et al., "Emotion recognition from eeg signal focusing on deep learning and shallow learning techniques," *IEEE Access*, vol. 9, 2021.

[142] T. Song et al., "Variational instance-adaptive graph for eeg emotion recognition," *IEEE TAC*, 2021.

[143] G. Zhao et al., "Multi-target positive emotion recognition from eeg signals," *IEEE TAC*, 2020.

[144] B. Hosseinifard et al., "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from eeg signal," *Computer methods and programs in biomedicine*, vol. 109, no. 3, pp. 339–345, 2013.

[145] J. Shen et al., "An improved empirical mode decomposition of electroencephalogram signals for depression detection," *IEEE TAC*, vol. 13, no. 1, pp. 262–271, 2019.

[146] M. Sharma et al., "An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with eeg signals," *Cognitive Systems Research*, vol. 52, pp. 508–520, 2018.

[147] H. Cai et al., "Study on feature selection methods for depression detection using three-electrode eeg data," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 10, pp. 558–565, 2018.

[148] S. Mahato et al., "Detection of major depressive disorder using linear and non-linear features from eeg signals," *Microsystem Technologies*, vol. 25, pp. 1065–1076, 2019.

[149] P. P. Thoduparambil et al., "Eeg-based deep learning model for the automatic detection of clinical depression," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 4, pp. 1349–1360, 2020.

[150] X. Li et al., "Eeg-based mild depression recognition using convolutional neural network," *Medical & biological engineering & computing*, vol. 57, no. 6, pp. 1341–1352, 2019.

[151] M. Kang et al., "Deep-asymmetry: Asymmetry matrix image for deep learning method in pre-screening depression," *Sensors*, vol. 20, no. 22, p. 6526, 2020.

[152] D. Wang et al., "Identification of depression with a semi-supervised gcn based on eeg data," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021, pp. 2338–2345.

[153] A. Saeedi et al., "Major depressive disorder diagnosis based on effective connectivity in eeg signals: A convolutional neural network and long short-term memory approach," *Cognitive Neurodynamics*, vol. 15, no. 2, pp. 239–252, 2021.

[154] G. Rodriguez-Bermudez et al., "Analysis of eeg signals using nonlinear dynamics and chaos: a review," *Applied mathematics & information sciences*, vol. 9, no. 5, p. 2309, 2015.

[155] S. A. Akar et al., "Nonlinear analysis of eegs of patients with major depression during different emotional states," *Computers in biology and medicine*, vol. 67, pp. 49–60, 2015.

[156] A. Ortiz et al., "The futility of long-term predictions in bipolar disorder: mood fluctuations are the result of deterministic chaotic processes," *International journal of bipolar disorders*, vol. 9, no. 1, pp. 1–7, 2021.

[157] A. Kustubayeva et al., "Lyapunov exponent of theta rhythm as a marker of depression during attentional network test," *Biological Psychiatry*, vol. 89, no. 9, p. S164, 2021.

[158] G. Chopra et al., "Using machine learning algorithms classified depressed patients and normal people," *International Journal of Machine Learning for Sustainable Development*, vol. 4, no. 1, pp. 31–40, 2022.

[159] L. Zhao et al., "Cardiorespiratory coupling analysis based on entropy and cross-entropy in distinguishing different depression stages," *Frontiers in physiology*, vol. 10, p. 359, 2019.

[160] M. Cukic et al., "Eeg machine learning with higuchi fractal dimension and sample entropy as features for successful detection of depression," *arXiv preprint arXiv:1803.05985*, 2018.

[161] L. Zhao et al., "Frontal alpha eeg asymmetry variation of depression patients assessed by entropy measures and lemple–ziv complexity," *JMBE*, vol. 41, no. 2, pp. 146–154, 2021.

[162] Y. Li et al., "Eeg-based mild depressive detection using differential evolution," *IEEE Access*, vol. 7, pp. 7814–7822, 2018.

[163] H. Cai et al., "Feature-level fusion approaches based on multimodal eeg data for depression recognition," *Information Fusion*, vol. 59, pp. 127–138, 2020.

[164] O. Faust et al., "Depression diagnosis support system based on eeg signal entropies," *Journal of mechanics in medicine and biology*, vol. 14, no. 03, 2014.

[165] W. Mumtaz et al., "A machine learning framework involving eeg-based functional connectivity to diagnose major depressive disorder (mdd),"

Medical & biological engineering & computing, vol. 56, pp. 233–246, 2018.
[166] M. Bachmann et al., "Methods for classifying depression in single channel eeg using linear and nonlinear signal analysis," *Computer methods and programs in biomedicine*, vol. 155, pp. 11–17, 2018.
[167] Z. Wan et al., "Hybrideegnet: A convolutional neural network for eeg feature learning and depression discrimination," *IEEE Access*, vol. 8, pp. 30 332–30 342, 2020.
[168] A. Saeedi et al., "Depression diagnosis and drug response prediction via recurrent neural networks and transformers utilizing eeg signals," *arXiv preprint arXiv:2303.06033*, 2023.
[169] X. Sun et al., "A novel complex network-based graph convolutional network in major depressive disorder detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–8, 2022.
[170] H. Wu and J. Liu, "A multi-stream deep learning model for eeg-based depression identification," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 2029–2034.
[171] H. Wu et al., "Eeg-based depression identification using a deep learning model," in *2022 IEEE 6th Conference on Information and Communication Technology (CICT)*. IEEE, 2022, pp. 1–5.
[172] B. M. Appelhans et al., "Heart rate variability as an index of regulated emotional responding," *Review of general psychology*, vol. 10, no. 3, pp. 229–240, 2006.
[173] H. R. Variability, "Standards of measurement, physiological interpretation, and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Eur Heart J*, vol. 17, no. 3, pp. 354–381, 1996.
[174] V. K. Jandackova et al., "Heart rate variability and depressive symptoms: a cross-lagged analysis over a 10-year period in the whitehall ii study," *Psychological medicine*, vol. 46, no. 10, pp. 2121–2131, 2016.
[175] B. Ay et al., "Automated depression detection using deep representation and sequence learning with eeg signals," *Journal of medical systems*, vol. 43, no. 7, pp. 1–12, 2019.
[176] S. Byun et al., "Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol," *Computers in biology and medicine*, vol. 112, p. 103381, 2019.
[177] T. Roh et al., "Wearable depression monitoring system with heart-rate variability," in *2014 36th EMBC*. IEEE, 2014, pp. 562–565.
[178] T. Matsui et al., "Impaired parasympathetic augmentation under relaxation in patients with depression as assessed by a novel non-contact microwave radar system," *Journal of Medical Engineering & Technology*, vol. 40, no. 1, pp. 15–19, 2016.
[179] G. Sun et al., "An objective screening method for major depressive disorder using logistic regression analysis of heart rate variability data obtained in a mental task paradigm," *Frontiers in Psychiatry*, vol. 7, p. 180, 2016.
[180] D. Kuang et al., "Depression recognition according to heart rate variability using bayesian networks," *Journal of psychiatric research*, vol. 95, pp. 282–287, 2017.
[181] S. T. Noor et al., "Predicting the risk of depression based on ecg using rnn," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.
[182] X. Zang et al., "End-to-end depression recognition based on a one-dimensional convolution neural network model using two-lead ecg signal," *JMBE*, vol. 42, no. 2, pp. 225–233, 2022.
[183] S. Mohanraj et al., "A deep convolution neural network framework for detecting depression," in *2022 6th ICICCS*. IEEE, 2022, pp. 1061–1068.
[184] R. Hudson et al., "Phytocannabinoids modulate emotional memory processing through interactions with the ventral hippocampus and mesolimbic dopamine system: implications for neuropsychiatric pathology," *Psychopharmacology*, vol. 235, pp. 447–458, 2018.
[185] J. Jiang et al., "Effect of multiple neonatal sevoflurane exposures on hippocampal apolipoprotein e levels and learning and memory abilities," *Pediatrics & Neonatology*, vol. 59, no. 2, pp. 154–160, 2018.
[186] M. A. O. Santos et al., "Global hippocampal atrophy in major depressive disorder: a meta-analysis of magnetic resonance imaging studies," *Trends in psychiatry and psychotherapy*, vol. 40, pp. 369–378, 2018.
[187] P. Chagué et al., "Radiological classification of dementia from anatomical mri assisted by machine learning-derived maps," *Journal of Neuroradiology*, vol. 48, no. 6, pp. 412–418, 2021.
[188] M. Mousavian et al., "Depression detection using feature extraction and deep learning from smri images," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. IEEE, 2019, pp. 1731–1736.
[189] J. X. Zheng et al., "Disrupted spontaneous neural activity related to cognitive impairment in postpartum women," *Frontiers in psychology*, vol. 9, p. 624, 2018.
[190] J. Zhao et al., "Functional network connectivity (fnc)-based generative adversarial network (gan) and its applications in classification of mental disorders," *Journal of Neuroscience Methods*, vol. 341, p. 108756, 2020.
[191] P. C. Mulders et al., "Resting-state functional connectivity in major depressive disorder: a review," *Neuroscience & Biobehavioral Reviews*, vol. 56, pp. 330–344, 2015.
[192] S. Yu et al., "Toward probabilistic diagnosis and understanding of depression based on functional mri data analysis with logistic group lasso," *PLOS ONE*, vol. 10, no. 5, pp. e0 123 524–, 2015.
[193] A. Yamashita et al., "Generalizable brain network markers of major depressive disorder across multiple imaging sites," *PLoS biology*, vol. 18, no. 12, p. e3000966, 2020.
[194] J. V. Haxby et al., "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, 2001.
[195] B. Yan et al., "Quantitative identification of major depression based on resting-state dynamic functional connectivity: A machine learning approach," *Frontiers in Neuroscience*, vol. 14, 2020.
[196] X. Wang et al., "Depression disorder classification of fmri data using sparse low-rank functional brain network and graph-based features," *Computational and mathematical methods in medicine*, vol. 2017, 2017.
[197] R. Bhaumik et al., "Multivariate pattern analysis strategies in detection of remitted major depressive disorder using resting state functional connectivity," *Neuroimage Clinical*, p. 390, 2016.
[198] A. Lord et al., "Changes in community structure of resting state functional connectivity in unipolar depression," 2012.
[199] G. E. Hinton et al., "Transforming auto-encoders," *Springer, Berlin, Heidelberg*, 2012.
[200] R. Gui et al., "The impact of emotional music on active roi in patients with depression based on deep learning: A task-state fmri study," *Computational Intelligence and Neuroscience*, vol. 2019, no. 6, pp. 1–14, 2019.
[201] D. Yao et al., "Triplet graph convolutional network for multi-scale analysis of functional connectivity using functional mri," in *International Workshop on GLMI*. Springer, 2019, pp. 70–78.
[202] S. I. Ktena et al., "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, 2018.
[203] E. Jun et al., "Identifying resting-state effective connectivity abnormalities in drug-nave major depressive disorder diagnosis via graph convolutional networks," *Human Brain Mapping*, vol. 41, no. 17, 2020.
[204] Q. Wang et al., "Function mri representation learning via self-supervised transformer for automated brain disorder analysis," in *International Workshop on MLMI*. Springer, 2022, pp. 1–10.
[205] Y. Cheng et al., "Classification algorithms for brain magnetic resonance imaging images of patients with end-stage renal disease and depression," *Contrast Media & Molecular Imaging*, vol. 2022, 2022.
[206] T. Ma et al., "Distinguishing bipolar depression from major depressive disorder using fnirs and deep neural network," *Progress In Electromagnetics Research*, vol. 169, pp. 73–86, 2020.
[207] Y. Zhu et al., "Classifying major depressive disorder using fnirs during motor rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 4, pp. 961–969, 2020.
[208] L. Fu et al., "Reduced prefrontal activation during the tower of london and verbal fluency task in patients with bipolar depression: A multi-channel nirs study," *Frontiers in Psychiatry*, vol. 9, pp. 214–, 2018.
[209] M. Kawano et al., "Correlation between frontal lobe oxy-hemoglobin and severity of depression assessed using near-infrared spectroscopy," *Journal of Affective Disorders*, vol. 205, pp. 154–158, 2016.
[210] A. Manelis et al., "The role of the right prefrontal cortex in recognition of facial emotional expressions in depressed individuals: fnirs study," *Journal of Affective Disorders*, vol. 258, pp. 151–158, 2019.
[211] L. Gu et al., "Classification of depressive disorder based on rs-fmri using multivariate pattern analysis with multiple features," in *2017 4th IAPR ACPR*, 2017, pp. 61–66.
[212] L. L. Zeng et al., "Unsupervised classification of major depression using functional connectivity mri," *Human Brain Mapping*, 2014.
[213] M. Mousavian et al., "Depression detection from smri and rs-fmri images using machine learning," *Journal of Intelligent Information Systems*, vol. 57, no. 2, pp. 395–418, 2021.
[214] H. Song et al., "Automatic depression discrimination on fnirs by using general linear model and svm," in *2014 7th International Conference on Biomedical Engineering and Informatics*. IEEE, 2014, pp. 278–282.

[215] Y. Zhu and M. Mehta, “Machine learning approach on frontal lobe activity to assess depression in adults: Implications for rehabilitation outcomes,” 2017, pp. 1–2.

[216] Q. Yu et al., “Gnn-based depression recognition using spatio-temporal information: A fnirs study,” *IEEE J BIOMED HEALTH*, vol. 26, no. 10, pp. 4925–4935, 2022.

[217] J. Chao et al., “fnirs evidence for distinguishing patients with major depression and healthy controls,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2211–2221, 2021.

[218] X. Wang et al., “Identifying psychological symptoms based on facial movements,” *Frontiers in Psychiatry*, vol. 11, p. 607890, 2020.

[219] J. M. Girard et al., “Social risk and depression: Evidence from manual and automatic facial expression analysis,” in *FG*, 2013.

[220] J. Li et al., “A novel study for mdd detection through task-elicited facial cues,” in *2018 IEEE International Conference on BIBM*, 2018.

[221] A. Mulay et al., “Automatic depression level detection through visual input,” in *2020 Fourth WorldS4*, 2020.

[222] J. Rottenberg et al., “Emotion context insensitivity in major depressive disorder.” *Journal of Abnormal Psychology*, vol. 114, no. 4, pp. 627–639, 2005.

[223] L. M. Bylsma et al., “A meta-analysis of emotional reactivity in major depressive disorder,” *Clinical psychology review*, vol. 28, no. 4, pp. 676–691, 2008.

[224] A. Pampouchidou et al., “Facial geometry and speech analysis for depression detection,” in *2017 39th EMBC*. IEEE, 2017, pp. 1433–1436.

[225] S. Alghowinem et al., “Head pose and movement analysis as an indicator of depression,” in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 2013, pp. 283–288.

[226] H. Meng et al., “Depression recognition based on dynamic facial and vocal expression features using partial least square regression,” in *Proceedings of the 3rd ACM international workshop on AVEC*, 2013, pp. 21–30.

[227] S. Alghowinem et al., “Eye movement analysis for depression detection,” in *ICIP*, 2013.

[228] J. F. Cohn et al., “Detecting depression from facial actions and vocal prosody,” in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009.

[229] A. Jan et al., “Automatic depression scale prediction using facial expression dynamics and regression,” in *International Workshop on AVEC*, 2014.

[230] A. Pampouchidou et al., “Depression assessment by fusing high and low level features from audio, video, and text,” in *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, 2016, pp. 27–34.

[231] W. Guo et al., “Deep neural networks for depression recognition based on 2d and 3d facial expressions under emotional stimulus tasks.” *Frontiers in neuroscience*, vol. 15, p. 609760, 2021.

[232] A. Pampouchidou et al., “Video-based depression detection using local curvelet binary patterns in pairwise orthogonal planes.” IEEE, 2016, pp. 3835–3838.

[233] N. Cummins et al., “Diagnosis of depression by behavioural signals: a multimodal approach,” in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, 2013, pp. 11–20.

[234] M. Nasir et al., “Multimodal and multiresolution depression detection from speech and facial landmark features,” in *Proceedings of the 6th international workshop on AVEC*, 2016, pp. 43–50.

[235] T. Baltrušaitis et al., “Openface: an open source facial behavior analysis toolkit,” in *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2016, pp. 1–10.

[236] F. Eyben et al., “Opensmile: the munich versatile and fast open-source audio feature extractor,” in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462.

[237] G. Littlewort et al., “The computer expression recognition toolbox (cert),” in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 2011, pp. 298–305.

[238] M. Al Jazaery et al., “Video-based depression level analysis by encoding deep spatiotemporal features,” *IEEE Transactions on Affective Computing*, vol. 12, no. 1, pp. 262–268, 2018.

[239] W. C. De Melo et al., “Depression detection based on deep distribution learning,” in *2019 IEEE international conference on image processing (ICIP)*. IEEE, 2019, pp. 4544–4548.

[240] X. Zhou et al., “Learning content-adaptive feature pooling for facial depression recognition in videos,” *Electronics Letters*, vol. 55, no. 11, pp. 648–650, 2019.

[241] L. Chao et al., “Multi task sequence learning for depression scale prediction from video,” in *ACII*. IEEE, 2015, pp. 526–531.

[242] Y. Zhu et al., “Automated depression diagnosis based on deep networks to encode facial appearance and dynamics,” *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 578–584, 2017.

[243] W. C. De Melo et al., “Encoding temporal information for automatic depression recognition from facial analysis,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1080–1084.

[244] A. Jan et al., “Artificial intelligent system for automatic depression level analysis through visual and vocal expressions,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 3, pp. 668–680, 2017.

[245] W. C. De Melo et al., “A deep multiscale spatiotemporal network for assessing depression from facial dynamics,” *IEEE transactions on affective computing*, vol. 13, no. 3, pp. 1581–1592, 2020.

[246] G. E. Hinton et al., “A fast learning algorithm for deep belief nets,” *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[247] R. Flores et al., “Temporal facial features for depression screening,” in *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, 2022, pp. 488–493.

[248] H. Kaya et al., “Eyes whisper depression: A cca based multimodal approach,” in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 961–964.

[249] L. Wen etal., “Automated depression diagnosis based on facial dynamic analysis and sparse coding,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1432–1441, 2015.

[250] A. Ray et al., “Multi-level attention network using text, audio and video for depression prediction,” in *Proceedings of the 9th international on AVEC and workshop*, 2019, pp. 81–88.

[251] Y. Wang et al., “Automatic depression detection via facial expressions using multiple instance learning,” in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1933–1936.

[252] S. A. Montgomery et al., “A new depression scale designed to be sensitive to change,” *The British journal of psychiatry*, vol. 134, no. 4, pp. 382–389, 1979.

[253] M. Asgari et al., “Inferring clinical depression from speech and spoken utterances,” in *2014 IEEE international workshop on Machine Learning for Signal Processing*, 2014, pp. 1–5.

[254] E. W. McGinnis et al., “Giving voice to vulnerable children: machine learning analysis of speech detects anxiety and depression in early childhood,” *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2294–2301, 2019.

[255] P. Wu et al., “Automatic depression recognition by intelligent speech signal processing: A systematic survey,” *CAAI Transactions on Intelligence Technology*, 2022.

[256] M. Tlachac et al., “Early mental health uncovering with short scripted and unscripted voice recordings,” in *Deep Learning Applications, Volume 4*. Springer, 2022, pp. 79–110.

[257] M. Rodrigues Makiuchi et al., “Multimodal fusion of bert-cnn and gated cnn representations for depression detection,” in *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, 2019, pp. 55–63.

[258] R. Flores et al., “Depression screening using deep learning on follow-up questions in clinical interviews,” in *2021 20th IEEE ICMLA*, 2021, pp. 595–600.

[259] K. Takakusaki, “Neurophysiology of gait: from the spinal cord to the frontal lobe,” *Movement Disorders*, vol. 28, no. 11, pp. 1483–1491, 2013.

[260] P. L. Sheridan et al., “The role of higher-level cognitive function in gait: executive dysfunction contributes to fall risk in alzheimer’s disease,” *Dementia and geriatric cognitive disorders*, vol. 24, no. 2, pp. 125–137, 2007.

[261] L. Sloman et al., “Mood, depressive illness and gait patterns,” *The Canadian Journal of Psychiatry*, vol. 32, no. 3, pp. 190–193, 1987.

[262] T. C. Brandler et al., “Depressive symptoms and gait dysfunction in the elderly,” *The American Journal of Geriatric Psychiatry*, vol. 20, no. 5, pp. 425–432, 2012.

[263] M. B. Van Iersel et al., “Quantitative gait analysis to detect gait disorders in geriatric patients with depression,” *J Am Geriatr Soc*, vol. 53, no. 8, pp. 1441–2, 2005.

[264] J. Michalak et al., “The effects of mindfulness-based cognitive therapy on depressive gait patterns,” *Journal of Cognitive and Behavioral Psychotherapies*, vol. 11, no. 1, pp. 13–27, 2011.

[265] S. Radovanović et al., "Gait characteristics in patients with major depression performing cognitive and motor tasks while walking," *Psychiatry research*, vol. 217, no. 1-2, pp. 39–46, 2014.

[266] A. S. Naidu et al., "Does dual-task gait differ in those with late-life depression versus mild cognitive impairment?" *The American Journal of Geriatric Psychiatry*, vol. 27, no. 1, pp. 62–72, 2019.

[267] J. Michalak et al., "Embodiment of sadness and depression—gait patterns associated with dysphoric mood," *Psychosomatic medicine*, vol. 71, no. 5, pp. 580–587, 2009.

[268] M. R. Lemke et al., "Spatiotemporal gait patterns during over ground locomotion in major depression compared with healthy controls," *Journal of psychiatric research*, vol. 34, no. 4-5, pp. 277–283, 2000.

[269] S. Xu et al., "Emotion recognition from gait analyses: Current research and future directions," *arXiv preprint arXiv:2003.11461*, 2020.

[270] F. Gholami et al., "A microsoft kinect-based point-of-care gait assessment framework for multiple sclerosis patients," *IEEE J BIOMED HEALTH*, vol. 21, no. 5, pp. 1376–1385, 2016.

[271] M. J. Nordin et al., "A survey of gait recognition based on skeleton model for human identification," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 12, no. 7, pp. 756–763, 2016.

[272] Y. Wang et al., "Gait recognition based on 3d skeleton joints captured by kinect," in *IEEE ICIP*, 2016, pp. 3151–3155.

[273] J. Han et al., "Individual recognition using gait energy image," *PAMI*, vol. 28, no. 2, pp. 316–322, 2005.

[274] X. Yang et al., "Gait recognition based on dynamic region analysis," *Signal Processing*, vol. 88, no. 9, pp. 2350–2356, 2008.

[275] S. Chancellor et al., "Quantifying and predicting mental illness severity in online pro-eating disorder communities," in *Proceedings of the 19th ACM CSCW*, 2016, pp. 1171–1184.

[276] T. Nguyen et al., "Affective and content analysis of online depression communities," *IEEE TAC*, vol. 5, no. 3, pp. 217–226, 2014.

[277] J. Wolohan et al., "Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with nlp," in *Proceedings of the first international workshop on language cognition and computational models*, 2018, pp. 11–21.

[278] M. Nadeem, "Identifying depression on twitter," *arXiv preprint arXiv:1607.07384*, 2016.

[279] E. E. Newell et al., "You sound so down: Capturing depressed affect through depressed language," *Journal of Language and Social Psychology*, vol. 37, no. 4, pp. 451–474, 2018.

[280] M. De Choudhury et al., "Predicting depression via social media," in *Seventh international AAAI conference on weblogs and social media*, 2013.

[281] F. Magami et al., "Automatic detection of depression from text data: A systematic literature review," in *XVI Brazilian Symposium on Information Systems*, 2020, pp. 1–8.

[282] A. Zirikly et al., "Explaining models of mental health via clinically grounded auxiliary tasks," in *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, 2022, pp. 30–39.

[283] A. Mallol-Ragolta et al., "A hierarchical attention network-based approach for depression detection from transcribed clinical interviews," 2019.

[284] T. Shen et al., "Cross-domain depression detection via harvesting social media." International Joint Conferences on Artificial Intelligence, 2018.

[285] S. G. Burdisso et al., "A text classification framework for simple and effective early depression detection over social media streams," *Expert Systems with Applications*, vol. 133, pp. 182–197, 2019.

[286] Z. Peng et al., "Multi-kernel svm based depression recognition using social media data," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 43–57, 2019.

[287] M. Tlachac et al., "Screening for depression with retrospectively harvested private versus public text," *IEEE journal of biomedical and health informatics*, vol. 24, no. 11, pp. 3326–3332, 2020.

[288] T. Liu et al., "The relationship between text message sentiment and self-reported depression," *Journal of affective disorders*, vol. 302, pp. 7–14, 2022.

[289] J. Meyerhoff et al., "Analyzing text message linguistic features: Do people with depression communicate differently with their close and non-close contacts?" *Behaviour Research and Therapy*, p. 104342, 2023.

[290] G. Shen et al., "Depression detection via harvesting social media: A multimodal dictionary learning solution." in *IJCAI*, 2017, pp. 3838–3844.

[291] D. Li et al., "Modeling spatiotemporal pattern of depressive symptoms caused by covid-19 using social media data mining," *International Journal of Environmental Research and Public Health*, vol. 17, no. 14, p. 4988, 2020.

[292] X. Hu et al., "Depression tendency detection model for weibo users based on bi-lstm," in *IEEE ICAICA*, 2021, pp. 785–790.

[293] H. Dinkel et al., "Text-based depression detection on sparse data," *arXiv preprint arXiv:1904.05154*, 2019.

[294] H. A. Schwartz et al., "Towards assessing changes in degree of depression through facebook," in *Proceedings of the workshop on CLPsych: from linguistic signal to clinical reality*, 2014, pp. 118–125.

[295] T. Pedersen, "Screening twitter users for depression and ptsd with lexical decision lists," in *Proceedings of the 2nd workshop on CLPsych: from linguistic signal to clinical reality*, 2015, pp. 46–53.

[296] H. Almeida et al., "Detecting early risk of depression from social media user-generated content." in *CLEF (working notes)*, 2017.

[297] C. Tung et al., "Analyzing depression tendency of web posts using an event-driven depression tendency warning model," *Artificial Intelligence in Medicine*, vol. 66, pp. 53–62, 2016.

[298] X. Wang et al., "A depression detection model based on sentiment analysis in micro-blog social network," in *Trends and Applications in Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, 2013, pp. 201–213.

[299] A. B. Shatte et al., "Social media markers to identify fathers at risk of postpartum depression," 2019.

[300] M. Islam et al., "Depression detection from social network data using machine learning techniques," *Health information science and systems*, vol. 6, no. 1, pp. 1–12, 2018.

[301] D. Mowery et al., "Feature studies to inform the classification of depressive symptoms from twitter data for population health," *arXiv preprint arXiv:1701.08229*, 2017.

[302] E. Fast et al., "Empath: Understanding topic signals in large-scale text," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2016, p. 4647–4657.

[303] A. G. Reece et al., "Instagram photos reveal predictive markers of depression," *EPJ Data Science*, vol. 6, no. 1, p. 15, 2017.

[304] D. M. Blei et al., "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[305] P. Resnik et al., "Beyond lda: exploring supervised topic modeling for depression-related language in twitter," in *Proceedings of the 2nd workshop on CLPsych: from linguistic signal to clinical reality*, 2015, pp. 99–107.

[306] L. Mariñelarena-Dondena et al., "Predicting depression: a comparative study of machine learning approaches based on language usage," 2017.

[307] A. H. Orabi et al., "Deep learning for depression detection of twitter users," in *Proceedings of the Fifth Workshop on CLPsych: From Keyboard to Clinic*, 2018, pp. 88–97.

[308] F. Sadeque et al., "Uarizona at the clef erisk 2017 pilot task: linear and recurrent models for early depression detection," in *CEUR workshop proceedings*, vol. 1866. NIH Public Access, 2017.

[309] J. Joshi etal., "Multimodal assistive technologies for depression diagnosis and monitoring," *Journal on Multimodal User Interfaces*, vol. 7, pp. 217–228, 2013.

[310] M. Bilalpur et al., "Multimodal feature selection for detecting mothers' depression in dyadic interactions with their adolescent offspring," in *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 2023, pp. 1–8.

[311] Y. Gong et al., "Topic modeling based multi-modal depression detection," in *Proceedings of the 7th annual workshop on AVEC*, 2017, pp. 69–76.

[312] M. Niu et al., "Hcag: A hierarchical context-aware graph attention model for depression detection," in *ICASSP*. IEEE, 2021, pp. 4235–4239.

[313] M. Kächele et al., "Fusion of audio-visual features using hierarchical classifier systems for the recognition of affective states and the state of depression," *depression*, vol. 1, no. 1, pp. 671–678, 2014.

[314] M. Niu et al., "Multimodal spatiotemporal representation for automatic depression level detection," *IEEE TAC*, 2020.

[315] H. Pérez Espinosa et al., "Fusing affective dimensions and audio-visual features from segmented video for depression recognition: Inaoe-buap's participation at avec'14 challenge," in *Proceedings of the 4th International Workshop on AVEC*, 2014, pp. 49–55.

[316] L. Yang et al, "Dcnn and dnn based multi-modal depression recognition," in *ACII*. IEEE, 2017, pp. 484–489.

[317] ——, "Integrating deep and shallow models for multi-modal depression analysis hybrid architectures," *IEEE TAC*, vol. 12, no. 1, pp. 239–253, 2018.

[318] S. Alghowinem et al., "Interpretation of depression detection models via feature selection methods," *IEEE transactions on affective computing*, vol. 14, no. 1, pp. 133–152, 2020.

[319] M. R. Morales et al., "Speech vs. text: A comparative analysis of features for depression detection systems," in *2016 IEEE spoken language technology workshop (SLT)*. IEEE, 2016, pp. 136–143.

[320] K. C. Fraser et al., "Detecting late-life depression in alzheimer's disease through analysis of speech and language," in *Proceedings of the Third Workshop on CLPsych*, 2016, pp. 1–11.

[321] T. Al Hanai et al., "Detecting depression with audio/text sequence modeling of interviews." in *Interspeech*, 2018, pp. 1716–1720.

[322] C. Lau et al., "Improving depression assessment with multi-task learning from speech and text information," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2021, pp. 449–453.

[323] G. Lam et al., "Context-aware deep learning for multi-modal depression detection," in *ICASSP*. IEEE, 2019, pp. 3946–3950.

[324] E. Toto et al., "Audibert: A deep transfer learning multimodal classification framework for depression screening," 2021, pp. 4145–4154.

[325] R. Flores et al., "Audiface: Multimodal deep learning for depression screening," in *Machine Learning for Healthcare Conference*. PMLR, 2022, pp. 609–630.

[326] J. R. Williamson et al., "Detecting depression using vocal, facial and semantic communication cues," in *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, 2016, pp. 11–18.

[327] L. Yang et al., "Multimodal measurement of depression using deep learning models," in *Proceedings of the 7th annual workshop on audio/visual emotion challenge*, 2017, pp. 53–59.

[328] S. Yin et al, "A multi-modal hierarchical recurrent neural network for depression detection," in *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, 2019, pp. 65–71.

[329] M. Morales et al., "A linguistically-informed fusion approach for multimodal depression detection," in *proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*, 2018, pp. 13–24.

[330] M. Rohanian et al., "Detecting depression with word-level multimodal fusion." in *INTERSPEECH*, 2019, pp. 1443–1447.

[331] S. A. Qureshi et al., "The verbal and non verbal signals of depression–combining acoustics, text and visuals for estimating depression level," *arXiv preprint arXiv:1904.07656*, 2019.

[332] J. Gong et al., "Common and distinct patterns of intrinsic brain activity alterations in major depression and bipolar disorder: voxel-based meta-analysis," *Translational psychiatry*, vol. 10, no. 1, p. 353, 2020.

[333] N. Patil et al., "Depression and associated alzheimer's disease," 2021.

[334] M. L. Radell et al., "Depression in post-traumatic stress disorder," *Reviews in the Neurosciences*, vol. 31, no. 7, pp. 703–722, 2020.

[335] N. J. Wiles et al., "Physical activity and depression in adolescents: cross-sectional findings from the alspac cohort," *Social psychiatry and psychiatric epidemiology*, vol. 47, pp. 1023–1033, 2012.

[336] T. Khumalo et al., "The relationship between locus of control and depression: A cross-sectional survey with university students in botswana," *South African Journal of Psychiatry*, vol. 25, 2019.

[337] K. Gbyl et al., "Cortical thickness following electroconvulsive therapy in patients with depression: a longitudinal mri study," *Acta Psychiatrica Scandinavica*, vol. 140, no. 3, pp. 205–216, 2019.

[338] J. Joshi et al., "Relative body parts movement for automatic depression analysis," in *2013 Humaine association conference on affective computing and intelligent interaction*. IEEE, 2013, pp. 492–497.

[339] S. Bhatia et al., "Automated measurement of head movement synchrony during dyadic depression severity interviews," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE, 2019, pp. 1–8.

[340] J. Binnewies et al., "Associations between depression, lifestyle and brain structure: A longitudinal mri study," *NeuroImage*, vol. 231, p. 117834, 2021.

[341] N. Cummins et al., "A review of depression and suicide risk assessment using speech analysis," *Speech communication*, vol. 71, pp. 10–49, 2015.

[342] H. Kaya et al., "Predicting depression and emotions in the cross-roads of cultures, para-linguistics, and non-linguistics," in *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, 2019, pp. 27–35.

[343] B. Stasak et al., "Breaking age barriers with automatic voice-based depression detection," *IEEE Pervasive Computing*, vol. 21, no. 2, pp. 10–19, 2022.

[344] M. Bilalpur et al., "Shap-based prediction of mother's history of depression to understand the influence on child behavior," in *Proceedings of the 25th International Conference on Multimodal Interaction*, 2023, pp. 537–544.