IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 July 2023, accepted 31 July 2023, date of publication 10 August 2023, date of current version 17 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3304236

SURVEY

Machine Learning in ADHD and Depression Mental Health Diagnosis: A Survey

CHRISTIAN NASH[1], **(Student Member, IEEE), RAJESH NAIR**[2], **AND SYED MOHSEN NAQVI**[1], **(Senior Member, IEEE)**

[1]Intelligent Sensing and Communications Group, School of Engineering, Newcastle University, NE1 7RU Newcastle upon Tyne, U.K.
[2]Adult ADHD Services, Cumbria, Northumberland, Tyne and Wear, NHS Foundation Trust, NE3 3XT Newcastle upon Tyne, U.K.

Corresponding authors: Christian Nash (c.nash@newcastle.ac.uk) and Syed Mohsen Naqvi (mohsen.naqvi@newcastle.ac.uk)

ABSTRACT This paper explores the current machine learning based methods used to identify Attention Deficit Hyperactivity Disorder (ADHD) and depression in humans. Prevalence of mental ADHD and depression is increasing worldwide, partly due to the devastating impact of the COVID-19 pandemic for the latter but also because of the increasing demand placed on the mental health services. It is known that depression is the most common mental health condition, affecting an estimated 19.7% of people aged over 16. ADHD is also a very prevalent mental health condition, affecting approximately 7.2% of all age groups, with this being conceived as a conservative estimate. We explore the use of machine learning to identify ADHD and depression using different wearable and non-wearable sensors/modalities for training and testing. These modalities include functional Magnetic Resonance Imagery (fMRI), Electroencephalography (EEG), Medical Notes, Video and Speech. With mental health awareness on the rise, it is necessary to survey the existing literature on ADHD and depression for a machine learning based reliable Artificial Intelligence (AI). With access to in-person clinics limited and a paradigm shift to remote consultations, there is a need for AI-based technology to support the healthcare bodies, particularly in developed countries.

INDEX TERMS Artificial intelligence, attention deficit hyperactivity disorder, depression, machine learning, mental health.

I. INTRODUCTION

There are a multitude of mental health conditions that can affect individuals, with various explanations accounting for their occurrence. There is no single definitive answer that has been identified. Conditions like depression and schizophrenia have been associated with hereditary factors and chemical imbalances in the human body [1]. However, this research mainly focuses on ADHD and depression, the two most prevalent mental disorders in humans. Both conditions often co-occur, with people diagnosed with one being more likely to be diagnosed with the other. In fact, adults with ADHD are three times more likely to have depression, and individuals with depression have a 30-40% prevalence of ADHD. There are also links between ADHD and increased suicidal ideation. Distinguishing between the two can be challenging due to overlapping symptoms and the potential side effects of ADHD medications. Saying this, differences exist in mood, motivation and sleep patterns between the two conditions [2], [3]. Both ADHD and depression are very broad topics, so to specialise our paper we focus only on wearable/non-wearable sensing and machine learning. Due to the link in symptoms, if one machine learning model can accurately detect one of the disorders, there is a chance that the model can be generalised in identifying the other. These connections and the high prevalence rate is what motivated this paper.

ADHD is a global concern affecting both children and adults. A 2015 meta-study found the worldwide prevalence of ADHD among children aged 18 and under to be 7.2% with a 95% confidence level [4]. Notably, cases of persistent ADHD, where symptoms that begin in childhood continue into adulthood, have a lower prevalence of 2.58% [5]. This discrepancy is believed to stem from limited access to diagnosis during

The associate editor coordinating the review of this manuscript and approving it for publication was Chih-Yu Hsu.

youth, suggesting that the real prevalence of ADHD in adults could be higher.

ADHD diagnosis is influenced by gender, with a male-to-female ratio of 2.28:1 observed in a sample of 858 ADHD diagnosed participants [6]. It's worth noting, however, that this ratio varies across different studies and regions, with a consistent trend of higher prevalence in males. An investigation into underdiagnosis in London found an undiagnosed ADHD rate of approximately 12% among 226 participants [7]. This underdiagnosis is often due to symptoms being misinterpreted as simple misbehaviour by parents and teachers. The British Broadcasting Company (BBC) suggests that the issue of undiagnosed ADHD is widespread. It estimates that around 1.5 million adults in the UK have ADHD, but only 120,000 are officially diagnosed [8]. In addition, those seeking a diagnosis may face substantial wait times of up to seven years [8].

The ramifications of ADHD extend beyond the individuals directly affected. It impacts families, with studies suggesting that an ADHD diagnosis can lead to higher divorce rates [9]. A longitudinal study by the University of Pittsburgh recorded a 22.7% divorce rate among families with ADHD, compared to 12.6% in non-ADHD families [10]. Moreover, ADHD carries significant economic implications. In the US, the annual cost of ADHD was estimated to range from 143−266 billion, with productivity-related adult income losses being the primary cost factor, accounting for 87−138 billion [11]. Additionally, a meta-analysis revealed a strong link between ADHD and criminal behavior, with individuals diagnosed with ADHD in childhood being two to three times more likely to be arrested, convicted, or incarcerated as adults [12].

The process of diagnosing ADHD can be intricate and lengthy, involving comprehensive history collection of an individual's behaviour across home and school environments [13]. However, several challenges can limit the effectiveness of this process, including variability in the subjective judgments made by assessors, inaccuracy or incompleteness of assessment questionnaires, and cultural considerations in the standardization of ADHD tests such as Conners-3 [14]. With these limitations, researchers are exploring alternative methods for diagnosis, including machine learning. Current techniques in ADHD diagnosis research often involve analyzing an individual's brain activity during specific tasks using fMRI and EEG. Such objective measurements, compared between ADHD patients and healthy controls, could offer significant insights. Moreover, further exploration of longitudinal studies could provide valuable knowledge about the cause and progression of ADHD.

Mental illnesses, with depression being the most prevalent, significantly impact the lives of those affected. As of 2014, it was estimated that nearly 19.7% of individuals aged 16 and above experienced symptoms of depression [15]. However, the World Health Organization reported that the diagnosed depression rates in the UK were a mere 4.5% as of 2015 [16]. This discrepancy suggests potential issues with the diagnostic process, a concern that further fuels the motivation for this work. Further complicating matters is the fact that 70-75% of people with diagnosable mental illnesses do not receive any treatment [17], [18]. The consequences of this treatment gap are significant, as evident in workplace-related mental health issues. From 2018 to 2019, stress and depression accounted for 44% of all work-related illness cases. Furthermore, it is estimated that up to 55% of all lost working days were due to mental health conditions [19]. These lost workdays bear a substantial economic cost, estimated between £74-99 billion [20].

The situation took a turn for the worse with the advent of the COVID-19 pandemic in 2020. A UK government report using the General Health Questionnaire 12 (GHQ-12) measure showed that average mental distress in April 2020 rose by 8.1% compared to the 2017 to 2019 average [21]. A study involving 1,300 healthcare providers and 6,200 non-healthcare providers showed that caregivers exhibited higher rates of depression, likely due to the harsh impact of the virus on UK care homes [22]. The pandemic's toll was also felt in mental health services. Data from South London services revealed that between March and June 2020, there were 1,109 additional deaths among their patients compared to previous years, with 64% of these fatalities attributed to COVID-19 [23]. Moreover, studies suggest that adults with mental health conditions were more likely to be hospitalised and even succumb to COVID-19 [24]. Given the gravity of these findings, the development and deployment of viable AI solutions for mental health detection are more urgent than ever.

There's a recognized strong interrelation between ADHD and depression, although the underlying causes remain elusive. Some theories propose that adults with ADHD are at an increased risk of experiencing adverse life events, which may contribute to the relationship between these two conditions. This hypothesis was tested in a study of 230 adults diagnosed with ADHD [25]. The data was processed using linear and logistic regression models, which revealed that individuals who had experienced adverse life events had a higher tendency towards depression. Further research supports the suggestion that an ADHD diagnosis may predispose individuals to develop depression in later life. This notion is backed by a longitudinal study that examined the data of 8310 children with ADHD and found an increased risk of recurrent depression in young adulthood. Furthermore, Mendelian randomization (MR) analyses have indicated a possible causal effect of ADHD genetic liability on major depression later in life [26]. The findings from these studies underscore the complex interplay between ADHD and depression, suggesting an urgent need for more focused research in this area.

Existing work in ADHD and depression analysis using machine learning methods has so far exploited either non-wearable data or wearable data. The most popular data sources to analyse for recognition of both ADHD and Depression are EEG signal data and MRI imaging data.

A vast range of machine learning methods have also been employed. Saying this, the most popular classification techniques are Support Vector Machines (SVM) and Neural Networks.

SVMs are a powerful supervised machine learning model primarily used for classification or regression tasks. SVMs work by identifying an optimal hyperplane that maximally separates different classes of data in a multi-dimensional space, effectively finding the decision boundary that has the largest margin between classes. Advantages of SVMs include their robustness in high-dimensional spaces, effectiveness when the number of dimensions exceeds the number of samples, and flexibility through the use of different kernel functions to capture complex decision boundaries. However, SVMs can be computationally intensive, especially for large datasets, they're less effective with noisy data where classes overlap, and they require proper tuning and selection of the kernel function and regularization parameter to perform optimally. The lack of a probabilistic interpretation of the results could also be seen as a disadvantage. Neural networks are a class of machine learning models inspired by the biological structure of the brain. They consist of interconnected layers of nodes or "neurons" that can learn to represent and manipulate data. Neural networks are particularly advantageous for their capacity to learn complex, non-linear relationships directly from raw data, making them useful for tasks like image recognition, natural language processing, and more. They can handle high-dimensional data and are highly scalable. However, neural networks also have some notable disadvantages. They require large amounts of labelled data for training, and they are often computationally expensive, both in terms of memory and processing power. The training process can also be challenging due to issues like overfitting, vanishing or exploding gradients. Lastly, the "black box" nature of neural networks can make the interpretation of their internal workings and decision processes difficult, posing challenges for transparency and trust.

Part of this work has been published at the International Conference on Information Fusion 2022 [27]. This is the complete version of the survey, extensively covering the vast majority of work completed in the area with broad explanations of engineering and medical techniques. Not all literature can be included due to page limitations. The authors would like to acknowledge the existence of surveys into detecting Mental Health using Machine Learning [28], [29], [30], [31], [32]. Saying this, they are different to this survey paper in several ways, with the absolute focus of our paper being ADHD and depression.

The rest of this paper is organised as follows. In Section II, the selection of literature is provided, with the parameters for acceptance being discussed. In Section III, testing for mental health conditions is presented for both ADHD and depression. Section IV provides insight into the publicly available datasets that are used in some of the studies analysed throughout this survey. Sections V and VI discuss the existing literature where machine learning has been exploited to diagnose ADHD and depression, respectively. Lastly, conclusions and future direction has been discussed in Section VII.

II. LITERATURE SELECTION CRITERIA

Before the paper compilation, research questions were proposed to allow for concise conclusions and efficient searches.

A. RESEARCH QUESTIONS (RQs)

The following research questions were finalized to focus the scope of the survey:

- What wearable and non-wearable sensing have been used in datasets for mental health ML-based ADHD and depression detection research?
- What are the advantages and disadvantages with individual modalities when trying to diagnose ADHD or depression using machine learning?
- What is the most popular classification algorithm applied?
- What is the standard of the classification method used?

B. SEARCH STRATEGY

Compiling the papers was achieved through a keyword string query over several literature databases. The keywords in the string were chosen to produce results that fit the RQs. The keyword string query is as follows: ("Classification" OR "Neural Networks" OR "Machine Learning" OR "Deep Learning" OR "Supervised Learning" OR "Unsupervised Learning") AND ("Depression" OR "ADHD") AND ("Diagnosis"). The following query was used on the following literature databases: IEEE Xplore, Science Direct, PubMed and Web of Science.

C. CRITERIA FOR IDENTIFICATION OF STUDIES

The inclusion criteria:

- Publication in English.
- Inclusion of data containing an individual with a formal diagnosis of depression or ADHD.
- Articles involving the diagnosis of a mental health condition by using Machine Learning.
- Investigating the diagnosis of ADHD or depression in humans using Machine Learning.
- Publication in a peer-reviewed journal.
- Publication within the last 11 years (2011-2022).

The exclusion criteria:

- Publication in a non-peer-reviewed journal.
- Publication in conference proceedings, book chapters and dissertations.

III. TESTING FOR A MENTAL HEALTH CONDITION

The DSM, currently in its fifth edition, serves as a widely utilized handbook for clinicians and psychiatrists in the United States [13]. The DSM-V encompasses the majority of mental health disorders and undergoes continuous professional revision. Over four hundred experts from thirteen different countries contributed to its development, representing fields such as epidemiology, neurology, paediatrics, primary care, psychiatry, psychology, and research methodology.

The DSM-V includes descriptions, symptoms, and other relevant criteria for specific mental health disorders to aid in diagnosis. Moreover, it provides diagnostic criteria for both children and adults. As a result, the majority of studies referenced in this paper employ the DSM-V to accurately identify individuals with ADHD or depression.

The International Classification of Diseases (ICD-11) was created by the World Health Organisation at a similar time to the DSM-V [33]. Similarly, it provides a broad range of knowledge on the extent, causes and consequences of human diseases (both medical and mental). The ICD-11 allows for systematic recording, interpretation and therefore analysis of mortality and morbidity data that is collected globally.

As both the DSM-V and ICD-11 are very similar in nature, there is a push to harmonise both together. To make this happen, in new iterations, the main focus will be to have the greatest clinical impact. Achieving this means increasing their international uniformity, with the enhancement of cultural compatibility being the primary goal.

A. ADHD

Regarding ADHD specifically, the DSM-V asserts that to be diagnosed with Attention Deficit Disorder (ADD), an individual must exhibit five or more symptoms of inattention persisting for over six months. Additionally, to be diagnosed with ADHD, five or more symptoms of both hyperactivity and impulsivity, along with inattention symptoms, must be present for more than six months.

Symptoms are classified into three major components: Inattention, Hyperactivity and Impulsivity [13], [34]:

- Inattention:
  1) Forgetfulness in daily tasks/work.
  2) Making careless mistakes in work or tasks.
  3) Difficulty sustaining attention in tasks.
  4) Fails to complete tasks.
  5) Doesn't listen when spoken directly to.
  6) Reluctance in joining tasks that require sustained attention.
  7) Often loses things necessary for tasks.
  8) Easily distracted by external stimuli.
  9) Often forgetful in daily tasks.
- Hyperactivity:
  1) An individual constantly moving around, even during inappropriate times such as in a cinema.
  2) Fidgeting with their hands excessively.
  3) Tapping surfaces with their fingers or tapping their feet on the ground.
  4) Excessive talking.
  5) Difficulty engaging in leisure activities quietly.
  6) Always on the go.
- Impulsivity:
  1) Interrupting conversations or answering before the question has been asked in full.
  2) Making a decision in the short term without considering the effects of the long term.
  3) Difficulty with self control.

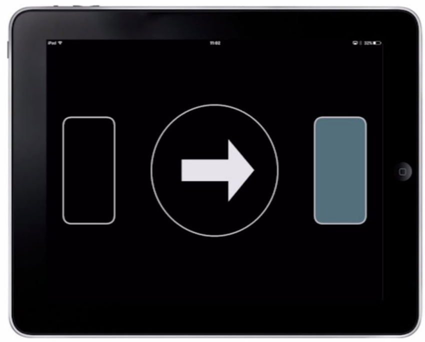

FIGURE 1. An example of the SST [107]. It can be done on a laptop or tablet, with the latter being recommended. The SST task takes roughly 15 minutes to complete. It is normal for both controls and ADHD subjects to be wrong 50% of the time.

1) QUESTIONNAIRES

Numerous questionnaires, including Conners-3 [14], can aid in the diagnosis of ADHD. These questionnaires can be completed by clinicians, patients, primary caregivers, or secondary caregivers. When completed by someone other than the clinician, the questionnaire offers valuable insight into the individual's behavioural history. However, the subjective nature of the responses may lead to inaccuracies and false positives in individuals pursuing a diagnosis, even if they do not genuinely have ADHD.

2) STOP SIGNAL TASK (SST)

Fig. 1 illustrates the screen a participant would encounter when taking the SST. This test represents a unique version of a classic method for measuring response inhibition (i.e., impulse control). Participants respond to an arrow stimulus by selecting one of two options based on the arrow's direction. The test comprises two parts:

- First, the participant is introduced to the test and instructed to press the left-hand button when they see a left-pointing arrow and the right-hand button when they see a right-pointing arrow. The participant practices this task in 16 trials.
- Next, the participant is asked to continue selecting buttons corresponding to the arrow directions. However, if an auditory signal (such as a beep) occurs, they should refrain from responding and not press the button.

3) CONTINUOUS PERFORMANCE TEST (CPT)

The CPT is a task-oriented, computerized assessment, that evaluates attention-related issues in individuals aged 8 years and older. It measures the participant's performance in areas such as attentiveness, impulsivity, sustained attention, and vigilance. The CPT supplements information obtained from rating scales like Conners-3 [14], offering insights into an individual's performance in attention tasks.

B. DEPRESSION

The DSM-V presents depression as persistent feelings of sadness and hopelessness while showing lack of interest in activities that were once enjoyed [13]. It mentions that individuals could experience additional physical symptoms such as chronic pain or digestive issues. The DSM-V states that the subject must be experiencing five or more of the following symptoms during the same 2 week period:

1) Depressed mood most of the day, experienced nearly every day.
2) Noticeable diminished interest or pleasure in all (or almost all) activities.
3) Experiencing significant weight loss or gain with a decrease or increase in appetite.
4) A reduction of physical movement and thoughts slowing down.
5) Fatigue or loss of energy nearly every day.
6) Feelings of worthlessness nearly every day.
7) Diminished ability to think or concentrate nearly every day.
8) Recurrent thoughts of death.

At least one of the symptoms should be either a depressed mood or loss of interest or please. It should be noted that to receive a diagnosis of depression, the symptoms must cause the subject clinically significant distress and impairment to everyday life.

1) PATIENT HEALTH QUESTIONNAIRE-9 (PHQ-9)

The PHQ-9 is a self-administered diagnostic tool used for criteria-based diagnosis of depression, as established by Kroenke, Spitzer and Williams [35]. The initial study involved 6,000 patients from various clinics. Criterion validity, which is predictive of outcomes, and construct validity, which assesses how well a test measures its intended subject, were determined against a mental health professional-led interview and the 20-item Short-Form General Health Survey respectively [36].

The PHQ-9, comprising of only nine questions, is based on the actual nine criteria for DSM-V depressive disorders diagnosis and can also indicate depressive symptom severity. Subjects respond based on their feelings and thoughts over the past 2 weeks as presented in Table 1. Clinicians interpret answers and scores to determine the presence and severity of depression (Table 2). Four or more ticks in the bold area suggest a depressive disorder.

2) BECK DEPRESSION INVENTORY (BDI-II)

The BDI-II is in its second iteration and is one of the most widely used instruments for detecting depression [37]. It is similar to the PHQ-9 with respect to it being a self-report questionnaire with it being designed to measure the severity of a subject's depression. It consists of 21 questions where each respective question has a list of four statements that are arranged in increasing severity. Each question is focused on a particular symptom of depression. Its second revision aligned its questions with the DSM-IV criteria by having the answers focused on the last 2 weeks upon taking the test.

FIGURE 2. A typical MRI Scanner [108]. MRI scanners are expensive pieces of equipment so they are constantly in use at hospitals for multiple needs. The bed moves in and out of the main scanner depending on what area of the human body is being imaged.

C. MAGNETIC RESONANCE IMAGING (MRI)

Magnetic Resonance Imaging (MRI) is a type of scan that uses powerful magnetic fields and radio waves to provide highly detailed images of the inside of the body. A scan can last between fifteen and ninety minutes, depending on the size of the area being scanned. The main advantage of MRI scanners is that they are harmless to the subject. A main downside to them is that they can be claustrophobic. There are several types of MRI scanner measurements with the focus of this survey being functional MRI (f-MRI) and resting-state MRI (rs-MRI). Initially developed to showcase regional/localized, time-varying changes in brain metabolism, f-MRI scanners have gained popularity due to their versatility with invasive and non-invasive techniques, good spatial resolution, and relatively low cost [38]. In the context of the studies mentioned throughout this paper, f-MRI's primary use is to observe increased neural activity by having a subject perform a task while in an MRI scanner, as shown in Fig. 2.

Two techniques exist for tracking neural activity: invasive and non-invasive sensing. Invasive sensing involves injecting the participant with a contrast agent to identify increased local cerebral blood flow and changes in oxygen concentration. In contrast, non-invasive sensing employs Arterial Spin Labeling (ASL), which suffers from reduced sensitivity, increased acquisition time, and higher sensitivity to body motion.

Typically, the subject in the f-MRI performs a task using visual, auditory, or other stimuli to induce cognitive states. A two-condition design is commonly used to identify activation. In the studies discussed in this paper, condition one is a resting state where the individual is instructed to sleep, providing an unstressed brain activity baseline known as

TABLE 1. The 9 questions a subject will answer to determine whether they suffer from depression and the level of depression [35].

Over the last 2 weeks, how often have you been bothered by any of the following problems? (use "✓" to indicate your answer)	Not at all	Several Days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things.	0	1	**2**	**3**
2. Feeling down, depressed or hopeless.	0	1	**2**	**3**
3. Trouble falling or staying asleep, or sleeping too much.	0	1	**2**	**3**
4. Feeling tired or having little energy.	0	1	**2**	**3**
5. Poor appetite or overeating.	0	1	**2**	**3**
6. Feeling bad about yourself or that you are a failure or have let yourself or your family down.	0	1	**2**	**3**
7. Trouble concentrating on things such as reading the newspaper or watching television.	0	1	**2**	**3**
8. Moving or speaking so slowly that other people could have noticed. Or the opposite, being fidgety or restless that you have been moving around a lot more than usual.	0	1	**2**	**3**
9. Thoughts that you would be better off dead or of hurting yourself.	0	**1**	**2**	**3**

TABLE 2. Interpretation of the total scores when evaluating depression severity using the PHQ-9 [35].

Total Score	Depression Severity
1-4	Minimal depression
5-9	Mild depression
10-14	Moderate depression
15-19	Moderately severe depression
20-27	Severe depression

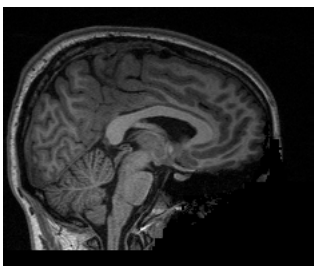

FIGURE 3. fMRI Scan Output from the ADHD-200 Dataset [39].

rs-MRI. Condition two involves the subject undertaking the designated activity. The images are combined to create a contrast map, and through image processing, an activation map is generated. Fig. 3 displays the output MRI image for a participant's resting state in the ADHD-200 dataset [39].

D. ELECTROENCEPHALOGRAPHY (EEG)

The brain consists of densely packed neurons interconnected through synapses, which serve as gateways for inhibitory or excitatory activity. Synaptic activity generates subtle electrical impulses, and when thousands of neurons fire in synchrony (due to localized brain area activation), an electrical field strong enough to penetrate tissue and the skull is produced. This enables humans to measure brain activity using specialized equipment, as shown in Fig. 4. The cap contains multiple electrodes that touch the individual's scalp to record brain activity, ensuring identical data collection between participants as the electrodes remain in fixed positions.

FIGURE 4. An advanced wearable EEG head cap [109]. The headcaps normally have a minimum of 8 electrodes with some advanced caps containing 256 electrodes. 32 electrodes have been found to be optimal for research purposes.

EEG offers a significant advantage as a non-invasive technique utilizing wearable sensors, providing highly accurate time-resolution data. This precise resolution enables researchers to detect activity within cortical areas on a sub-second timescale. As the voltage fluctuations measured at the

TABLE 3. **Different frequency bands and their purposes in an EEG machine.**

Name	Frequency Band (Hz)	Purpose
Delta	1 - 4	Used to assess the depth of sleep. The stronger the rhythm, the deeper the sleep. Research shows it to be associated with an increased concentration on internal memory tasks.
Theta	4 - 7	Associated with memory encoding and retrieval. Theta waves have strong inference with cognitive load.
Alpha	7 - 12	Alpha levels are increased when in a relaxed state.
Beta	12 - 30	Beta frequencies become more prominent when we plan to execute a body part.
Gamma	30 - 50	Researcher found discrepancies with the Gamma frequency. Some argue it reflects attentive focusing while others argue it represents rapid eye movement.

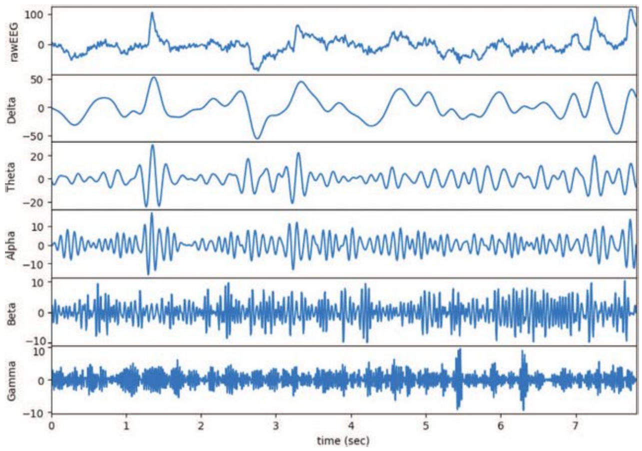

FIGURE 5. **Example of the raw EEG signal with the extracted common frequency bands [110].**

electrodes are small, the data is amplified to be displayed as a sequence of voltage values. Moreover, since the electrodes analyze brain activity across various regions, data can be examined in specific brain cortices. Different cortices within the brain are responsible for distinct activities, allowing certain tasks to isolate particular brain areas. The EEG identifies specific frequency bands, as illustrated in Table 3:

Fig. 5 is showing the output of wearing the EEG hat. It is showing the outputs in the form of the main frequency bands highlighted in Table 3.

E. AUDIO/VISUAL

Before recording a dataset that will be evaluated using machine learning, consideration of the equipment is vital. For instance, speech signals can be severely corrupted by background noise, depending on the recording environment. Therefore, it is recommended that a dual microphone configuration is implemented. Ideally, a microphone, such as a lavalier microphone, is attached to the participant being recorded. While a secondary microphone, or microphone array is placed in the room to record the environment noises. To achieve the best quality audio recording, it is recommended to record in the highest sampling rate that the chosen microphone has to offer. Pair the highest sampling rate with a 24-bit rate to increase the quality of the recordings while increasing the level of detail. When designing video data recording, it is vital to consider the stability of the cameras. Correct tripods and mounts are essential as you don't want any additional blurriness or motion captured. Depending on what is being captured, the resolution of the camera and frame rate can differ due to there being a trade-off between resolution and performance with machine learning algorithms. Provided the camera is stable, recording at a resolution of 1920×1080 (High-Definition) at a frame-per-second (fps) of 30 is suitable. If a budget allows for it, there is also the option to record in 4K (4096×2160) at 30 fps.

IV. DATASETS

There are not a lot of publicly available datasets for Mental Health challenges due to the highly sensitive nature of the data. The main concern is the protection of the participant's privacy i.e, identity and health information. Therefore in some cases, it is safer to not release the data publicly. In the cases where data has been made public, mainly the video modality, it is processed into features that can not be reverse transformed into their original format. This pre-processing of the data can impact the algorithms used to experiment with such data.

It is known that a common problem with datasets involving medical information are usually small in size. This is due to complications with preserving the participant's identity while also facing challenges in finding enough individuals with the condition being researched. If a researcher has access to the original video, the small dataset size could be increased using data augmentation techniques. The choice is with the researcher but as a few examples, blurring can be applied to the video or frame mirroring could be applied [40].

A. ADHD

1) ADHD-200

The ADHD-200 Machine Learning competition invited the neuroimaging and data mining communities to develop a pattern classification method that could distinguish brain activity differences between a control and an individual with ADHD [39]. The dataset comprises a combination of structural MRIs (s-MRIs) and resting state functional MRIs (rs-fMRI). The ADHD-200 dataset features pre-processed rs-fMRI data from 973 participants, categorized as Typically Developing (TD), ADHD-Impulsive (ADHD-I), and ADHD-Combined (ADHD-C). Table 4 displays the breakdown of the sample population. The training set released for the developed models contained 776 participants data. For testing the models, a further 197 data entries were released.

TABLE 4. The complete breakdown of training and testing data combined for the ADHD-200 dataset. Where PU is Peking University, UPitt is the University of Pittsburgh, NYU is the New York University Child Study Center, BrownU is the Bradley Hospital at Brown University, NI is NeuroIMAGE, OHSU is the Oregon Health and Science University, KKI is the Kennedy Krieger Institute and WashU is Washington University in St. Louis. There are a total of 973 participants in the dataset.

Site	Age	Male	Female	TD	ADHD	Total
PU	8 - 17	174	71	143	102	245
UPitt	10 -20	53	45	94	4	98
NYU	7 - 18	173	90	111	152	263
BrownU	8 - 18	9	17	26	0	26
NI	11 - 22	43	30	37	36	73
OHSU	7 - 121	60	53	70	43	113
KKI	8 - 13	56	38	69	25	94
WashU	7 - 22	33	28	61	0	61
Total	-	601	372	611	362	973

The dataset contains more controls than ADHD patients because the competition's primary goal was to accurately identify controls, while ADHD individual identification was a secondary objective. For each participant, the resting state fMRI data was processed through respective pipelines based on the Statistical Parametric Mapping 8 (SPM8) fMRI analysis package. The processing steps included:

1) Six parameter rigid body motion correction.
2) Non-linear spatial warping (involves estimating and interpolating) of each participant's anatomical volume to the MNMI T1 template space at a 1 × 1 × 1 mm resolution.
3) Interpolation of fMRI volumes into the T1 template space at a 3 × 3 × 3 mm spatial resolution.
4) Eight millimetre full width at half maximum (FWHM) Gaussian spatial filtering of fMRI volumes.
5) Truncation of resting state fMRI data to length 185 seconds and temporal linear interpolation of all scans into a sampling rate of 2Hz,

As a result of the pre-processing, all participants have resting state fMRI data that:

- is aligned in the MNI T1 Template space.
- have the same spatial dimensions of 57 × 67 × 50 voxels.
- have the same spatial resolution of 3 × 3 × 3 mm voxel size.
- have the same temporal dimensions of 370 time-points with a 0.5s volume time.

where a voxel is a 3-Dimensional unit of the image with a single value.

B. DEPRESSION

When the Audio/Visual Emotion Challenge and Workshop (AVEC) is held, there is often a state-of-mind detection challenge or a detecting depression challenge. The data is released to authors and a competition begins where authors can improve on the baselines and provide a state-of-the-art system. Normally datasets used are extensions of the previous challenges datasets. The last AVEC challenge was its 9^{th} proceedings in 2019 [41]. The vast majority of the work for depression evaluate their machine learning models on the AVEC2013 and AVEC2014 datasets [42], [43]. Their baseline scores are shown in Table 5.

TABLE 5. Showing the baseline Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for the AVEC2013 [42] and AVEC2014 [43] challenges.

	MAE	RMSE
2013	10.88	13.61
2014	8.86	10.86

1) DAIC-WOZ

The Distress Assessment Interview Corpus Wizard of Oz (DAIC-WOZ) was first introduced by Scherer et al [44]. Wizard of Oz interviews are conducted by having an animated interviewer (Ellie) who is controlled by a human interviewer in another room.

The DAIC-WOZ dataset contains clinical interviews designed to support the diagnosis of psychological distress conditions. It is composed of recordings and transcripts taken from 142 subjects that went through interviews with a computer agent. The computer agent, named Ellie, is the fundamental feature of a Wizard-Of-Oz style interview. Ellie is controlled by a human interviewer in another room with her function being to eradicate interviewer effects. The interviewer effect is the influence of the characteristics of an interviewer on the responses of the interviewee. To choose the subjects, recording took place over two sites, a United States (US) Veteran centre and the University of South California (USC) Institute for Creative Technologies. The interviews were conducted by one of two female interviewers, both having basic clinical experience.

The dataset is ever expanding but to begin with, there were 142 subjects with data being collected for: PHQ-8 score, gaze direction, pose, facial expressivity and acoustic indications [45]. The PHQ-8 score is a measure of the severity of the subject's depression. Along with the PHQ-8 score is a binary label for whether the subject has depression. Furthermore, the dataset is split into a training set consisting of 107 subjects where 30 are depressed and 77 are controls. There is then a validation set consisting of 35 subjects where 12 are depressed and 23 are controls.

V. MACHINE LEARNING AND COMPUTER VISION IN ADHD DETECTION

A. NON-WEARABLE TECHNIQUES

1) IMAGING

Table 6 shows that numerous studies have analyzed imaging data using various techniques. Exploiting an SVM is a popular approach for classification, applied to both imaging and EEG data. This popularity could be attributed to SVM's ability to capture complex relationships in the data or its resilience to overfitting. With increasing mental health awareness, it is anticipated that more advanced analysis methods,

TABLE 6. Summary of the research conducted using the ADHD-200 dataset. Where '-' denotes the authors have not specified the result or if validation has not been performed. To the best of the authors' knowledge, the official training and testing data split was used throughout the papers discussed in the table.

Classification Technique	Methodology	Accuracy (%)	Sensitivity (%)	Specificity (%)	Cross-Validation
Convolutional Neural Network (CNN)	Feature Extraction with Seed-based Correlation [111]	85.36	72.80	66.24	✓
	EM-MI Multi-instance Learning Algorithm [112]	70.40	-	-	✓
	4D-CNN Sigmoid Classifier [113]	71.30	73.20	69.70	-
	Two 3D-CNNs: One for sMRI data and the other for fMRI Data with the Result of Both Summed Together [114]	72.89	-	-	-
	Low-level Feature Extraction with Prior Knowledge [115]	69.15	-	-	✓
	Softmax classifiers [116]	73.10	65.50	91.60	-
	Dilated 3D-CNN [117]	76.66	39.00	89.10	✓
SVM	Functional Connectivity Networks with Multi-Dimensional Scaling [118]	73.55	75.00	72.73	✓
	Discriminate Subnetwork Selection [119]	94.91	93.22	96.94	✓
	Feature Selection with Functional Connectivity Matrices using Neural Networks [120]	73.30	-	-	✓
	R-RELIEF Algorithm [121]	81.82	33.34	100	-
	Stockwell Transform Followed by Fuzzy Entropy [47]	96.68	-	-	✓
	Bi-objective Optimisation [122]	81.08	-	-	✓
Extreme Learning Machine	F-score and Sequential Forward Selection [123]	90.18	-	-	✓
	Fast Independent Component Analysis [124]	98.20	70.80	-	-
Convolutional-GRU	Spatio-temporal Feature Extraction with Nested Residual Convolutional Denoising Autoencoding [125]	72.44	70.15	74.16	✓
k-Nearest Neighbour (kNN)	J-Extended Frobenius Norms [126]	81.00	66.00	87.00	-
Multi-Layer Perceptron (MLP)	Feature Extraction using Kernel Granger Causality and Correlation between Probabilities of Reoccurrence [127]	90.00 - 95.00	-	-	✓
MRI HoG-feature based patient classification	Histogram of Oriented Gradients [128]	69.60	79.80	57.10	-
Decision Tree	Non-negative Matrix Factorising [129]	66.80	76.20	50.60	✓
Random Forest	3D fMRI Segmentation [130]	75.46	-	-	✓
Projection Based Learning algorithm for a Meta-cognitive RBFNetwork	Discriminant Spatial Filtering [131]	73.83	68.15	81.50	✓
Deep Belief Network	Greedy Training Algorithm [132]	69.83	75.00	65.00	-
Linear Regression	Spatiotemporal Attention Autoencoder [133]	75.80	79.50	73.30	-
Binary classification based on hypothesis energy differences	$L_{2,1}$-Norm Linear Discrimination Analysis [46]	97.60	96.10	98.48	✓

such as deep learning, will become widespread, as summarized in Table 6. However, deep learning requires a large amount of data, and despite the ADHD-200 dataset being extensive, it is argued that without data augmentation, a deep learning algorithm cannot reliably train and predict outcomes. Researchers often used only partial amounts of the ADHD-200 dataset in their experiments to reduce bias within the models, as the dataset contains more controls than patients, causing models to learn control characteristics more effectively than those of ADHD patients.

Tang et al. achieved the best-performing study, obtaining a 97.6% testing accuracy with high sensitivity and specificity [46]. The selected ADHD features were brain Functional Connectivity (FC), processed through an $L_{2,1}$-Norm Linear Discrimination Analysis model. The model's output was then processed through a binary hypothesis testing framework to make a decision. The authors suggest that the binary hypothesis testing framework can alleviate some issues when testing with a smaller dataset. Among the studies incorporating an SVM, Sartipi et al. achieved the highest accuracy at 96.68% [47]. As fMRI data in different brain cortices change rapidly, a robust time-frequency transform (Stockwell Transformation) was applied. The time-frequency domains were then partitioned into sub-matrices for calculating their fuzzy entropies.

2) MEDICAL NOTES

Bledsoe et al. [48] utilized a SVM to achieve impressive classification accuracy, sensitivity, and specificity, each scoring a perfect 100%. However, it is crucial to consider the limitations of this study. The dataset size was quite small, and there is a substantial risk of overfitting. These results, while promising, may not generalize well to larger, more heterogeneous populations.

Subsequently, Chen et al. [49] and Tachmazidis et al. [50] both used a larger dataset made available by a National Health Service specialist mental health provider. While the former used a Decision Tree and achieved an accuracy of 85.51% and an AUC of 0.87, the latter used a hybrid AI model, achieving an accuracy of 95.7% for a three-way classification problem (ADHD/No-ADHD/Expert). These studies showcased the potential of advanced machine learning techniques in ADHD diagnosis while also indicating the importance of large, high-quality datasets in training effective models.

Further extending this approach, Christiansen et al. [51] applied LightGBM, a more advanced machine learning algorithm, to a multi-class classification problem. This problem involved differentiating not just between ADHD and control groups but also between subjects with obesity and problematic gambling. The algorithm achieved a global accuracy of 80%, with precision and recall varying between different classes, hinting at the potential complexity when more conditions are included.

Duda et al. [52] took a different approach by trying to differentiate between ADHD and Autism Spectrum Disorder (ASD) using machine learning. They found that classification performance fluctuated significantly depending on the dataset used, and the model that incorporated all data sources and utilized repeated cross-validation achieved an AUC of 0.89 ± 0.01. This study underlined the importance of data source selection and rigorous validation techniques, highlighting that these tools could serve as a promising new avenue for ADHD detection.

In summary, these studies indicate that machine learning holds significant promise in the field of ADHD diagnosis and differentiation from other mental health conditions. However, the robustness and generalizability of the findings are crucial, and careful attention needs to be paid to the dataset size, quality, and the potential for overfitting in model development. It also suggests that advanced algorithms like LightGBM and hybrid AI models can handle complex multi-class problems more effectively.

B. WEARABLE TECHNIQUES

Convolutional Neural Networks (CNNs) have been extensively applied in these studies. Amado-Caballero et al. [53] achieved a striking classification accuracy of 98.57%, as well as high sensitivity and specificity, demonstrating the potential of CNNs in interpreting EEG data. Chen et al. [54] and Moghaddari, Lighvan, and Danishvar [55] further validated this approach, with the latter achieving an impressive average testing accuracy of 98.48%. Ahmadi et al. [56] developed a sophisticated deep CNN model, extracting both spatial and frequency band features and yielding a nearly perfect classification accuracy of 99.46% for ADHD subtypes.

Simultaneously, studies using SVMs demonstrate that effective pre-processing techniques and SVM parameter selection can also yield high classification accuracies. Chang et al. [57] utilized a decorrelation method with an independent 2-sample t-test to achieve 80% accuracy, while Chen et al. [58] leveraged Mutual Information analysis to reach 85.7% accuracy. De-Dea et al. [59] and Rezaeezadeh, Shamekhi, and Shamsi [60] further substantiate the SVM's utility, with the latter achieving a remarkable 99.58% classification accuracy.

Additionally, wavelet transform techniques, when coupled with diverse classifiers, have shown promise in ADHD classification tasks. Notably, Tor et al. [61] achieved a 97.88% classification accuracy using a kNN classifier, demonstrating the potential efficacy of this technique.

Tenev et al. [62] took a unique approach, applying multiple classifiers to EEG data under various conditions to categorize ADHD subtypes and controls. While the accuracy fluctuated across different conditions (69.2%-82.3%), this approach underlines the importance of considering the context in which EEG data is collected and its impact on model performance.

Lastly, Poil et al. [63] emphasized the need to consider age and frequency effects on ADHD-related EEG signal alterations, pointing out an important factor for future research.

In conclusion, the analysis of EEG data using machine learning, particularly CNN and SVM models, has proved to be a promising tool for ADHD diagnosis and differentiation from other conditions. The studies reviewed indicate that careful pre-processing of EEG data and appropriate selection of model parameters are essential to achieve high classification accuracy. Furthermore, they highlight the importance of accounting for factors such as age, ADHD subtypes, and context in future research to ensure generalizability and clinical relevance of these models.

VI. MACHINE LEARNING AND COMPUTER VISION IN DEPRESSION DETECTION

A. NON-WEARABLE TECHNIQUES

1) VIDEO

Research by Zhu et al. [64] and Meshram and Rambola [65] use CNNs to extract facial features, either static or dynamic, to estimate the BDI-II depression severity score. The model by Zhu et al. notably improved upon results from the AVEC-2013 dataset, while Meshram and Rambola's model achieved a high classification average of 92.56% on the AVEC-2016 challenge dataset, suggesting the effectiveness of deep learning techniques in detecting depression.

Similar trends can be observed in the work of He et al. [66], who also employed the extraction of dynamic facial features, specifically using the MRLBP-TOP technique. The promising RMSE value of 8.90 achieved by this framework compared to the AVEC-2013 baseline (10.72) demonstrates the potential of advanced feature extraction techniques in this field.

Li et al. [67] took a slightly different approach by focusing on eye movement as a predictor of depression. Despite this novel approach, the model still achieved a classification accuracy of 80.1%, indicating that diverse biological signals can potentially be valuable in depression detection.

In the research conducted by Hong et al. [68], the focus was on distinguishing between bipolar and unipolar disorders, as well as identifying healthy controls. Their methodology combined action unit descriptors and motion vectors processed by machine learning techniques, yielding a reasonable classification accuracy of 72.2%.

Zhou et al. [69] introduced the MR-DepressNet, a deep regression network that used visual features to estimate depression severity. This model improved on the AVEC-2013 baseline RMSE value (13.61) by achieving 8.28, demonstrating the potential of visual feature exploitation in improving depression estimation.

The research of Tadalagi and Joshi [70], Shang et al. [71], and Uddin, Joolee and Lee [72] all took innovative approaches by combining various techniques and methodologies. Tadalagi and Joshi implemented their model on a real-time system, while Shang et al. introduced a quaternion-based method. Uddin, Joolee and Lee used a two-stream deep spatiotemporal network. All of these methods were evaluated on AVEC datasets and achieved competitive results.

Song et al. [73] focused on the extraction of multi-scale video-level features, providing a novel perspective on depression analysis. Using spectral representations processed by CNNs and Artificial Neural Networks (ANN), their model achieved a competitive MAE/RMSE score on the AVEC2013 test set.

Yang et al. [74] introduced a multi-modal framework that exploited video, audio and text data to estimate PHQ-8 scores and infer the mental condition of the subject. The combination of a deep CNN, DNN, Paragraph Vector, SVM, and random forest methods showed the value of integrating different types of data in depression detection.

Lastly, De Melo, Granger and López [75] tackled the cost-effectiveness of 3D-CNNs by proposing a deep learning architecture that operates without 3D convolutions. Despite the reduction in trainable parameters, their MDN still improved performance when compared to existing 3D ResNet models, illustrating the importance of resource optimization without compromising performance.

In conclusion, the research reviewed supports the notion that a multi-modal, multi-feature approach combining various machine learning techniques can effectively predict depression severity scores. The findings are congruent in that they all indicate the potential of machine learning and feature extraction in improving depression detection and analysis. However, they also highlight the complexity and the need for further research to enhance accuracy and achieve effective real-time applications.

2) AUDIO

The detection and diagnosis of depression have been significantly advanced by various machine learning approaches, particularly with the use of speech analysis. A selection of studies provides an intriguing narrative of the progress and development in this area, highlighting the power of speech as a rich source of information for depression analysis.

Arevian et al. [76] pioneered the use of lexical content, complexity, and vocal expression in tracking depression. They also leveraged latent semantic analysis (LSA), taking into account semantic coherence, which was an innovative step in expanding the feature pool. Their use of the SVM classifier established a fundamental framework for future research.

Building upon the traditional feature extraction methods, Chen et al. [77] delved deeper into the hidden structures within the feature set. Their introduction of a sparse stacked autoencoder to learn higher-quality deep features marked an important milestone in enhancing the efficacy of depression detection. The significant accuracy of 89% attained by this study, using an SVM for classification, underscored the need for high-quality features.

Meanwhile, Cummins et al. [78] innovated in a different direction, examining how depression impacts acoustic models of spectral features. By using Monte Carlo sampling, they introduced a robust feature estimation method that uniquely connected depression to the Acoustic Volume.

Taking deep learning to the forefront, He and Cao [79] used CNNs to extract deep-learned features from raw speech waveforms and spectrograms. This shift towards automated depression analysis tools that could generate complex features signaled a significant step towards more sophisticated models.

Conversely, Jiang et al. [80] proposed an ensemble logistic regression model, offering separate models for males and females. This gender-specific approach, combined with diverse feature extraction, added a new dimension to the field, recognizing the potential differences in depression expression across genders.

Meanwhile, Li et al. [81] introduced the Multiscale Audio Data Normalization (MADN) algorithm, marking another significant advancement in feature extraction. Their approach further emphasized the potential for innovative methods to improve upon existing models.

Pushing the boundaries of feature analysis, Muzammel et al. [82] focused on the acoustic features of vowel and consonant spaces. Their method of augmenting data and segmenting speech brought attention to the nuanced elements of speech and their potential role in depression detection.

A notable turning point occurs with Zhao et al. [83], who introduced a comprehensive approach combining unsupervised learning, hierarchical attention, and knowledge transfer. Their impressive results underscored the potential of complex deep learning architectures in determining depression severity.

In summary, this collection of studies forms an engaging narrative that underscores the evolution of speech analysis in depression detection. From basic lexical analysis to deep learning, and from gender-specific models to nuanced vowel and consonant analysis, it shows a consistent progression towards more complex and sophisticated models. This narrative serves as a testament to the ongoing advancement of machine learning techniques in mental health research, particularly in understanding and addressing depression.

3) IMAGING

Depression detection research has expanded beyond speech analysis to leverage advanced neuroimaging technologies, and machine learning has remained a constant ally. Various studies have illustrated how different feature extraction and classification methods can yield significant results in identifying depression at an individual level.

In a novel approach, Cao et al. [84] made strides by focusing on the individual rather than group dynamics, using probability density functions (PDFs) to target functional connectivity. They integrated a t-test for primitive selection, followed by Kernel density estimation for PDFs, resulting in a significant classification accuracy of 84.21% using an SVM classifier. This shift towards an individual-oriented approach marked a significant milestone in depression detection.

Building upon functional connectivity, Guo et al. [85] further demonstrated the potential of brain network analysis by using nonparametric permutation tests for group comparisons. The effective use of topological metrics as inputs to classifiers, particularly the SVM-RBF, yielded an impressive classification accuracy of 83.0%. Guo et al. [86] further refined their approach by constructing an automatic classifier based on a high-order minimum spanning tree functional brain network. Their multi-kernel SVM, after intricate feature extraction and selection, achieved an exceptional accuracy of 97.54%.

In parallel, Li et al. [87] used voxel-based morphometry (VBM) and regional homogeneity (ReHo) analyses to extract key features. By adopting the LASSO approach to isolate the most informative brain regions, they achieved a validated classification accuracy of 86.4% with an SVM classifier. Their emphasis on feature extraction highlighted the importance of selecting relevant regions in the brain for effective depression detection.

Additionally, Rosa et al. [88] proposed a sparse framework for depression classification. Their utilization of sparse inverse covariance models to estimate functional connectivity, coupled with an L1-norm SVM, resulted in an accuracy of 85%. This approach once again demonstrated the importance of functional connectivity in the detection of depression.

Conversely Li et al. [89], who leveraged independent component analysis to define the triple network model. Their integration of effective connectivity features, dynamic functional connectivity features, and rigorous statistical testing led to an accuracy of 90.91% with an SVM classifier.

Simultaneously, Sen et al. [90] focused on dynamic and static connectivity measures, extracted from rs-fMRI data, as a basis for feature extraction. Their use of Pearson's correlation and entropy measures resulted in a combination of static and dynamic features, yielding a classification result of 82% with an RBF-SVM.

Finally, Wang et al. [91] distinguished themselves by using functional near-infrared spectroscopy (fNIR) instead of fMRI. By utilizing the unique properties of near-infrared light and the absorptive characteristics of blood, they were able to extract crucial features. With an AlexNet structured network, they achieved an impressive accuracy of 90%.

Together, these studies form a compelling narrative that showcases the interplay between neuroimaging technologies and machine learning in depression detection. From functional connectivity and brain network analysis to innovative uses of light in fNIR, the research direction showcases a continued progression toward individual-level analysis and a growing emphasis on sophisticated feature extraction and selection methods. The consistent use of SVM classifiers across most studies points to their effectiveness in this context, further highlighting the importance of machine learning in mental health research.

4) MEDICAL NOTES

De-Souza et al. [92] sought to develop a machine learning tool to detect depression using an amalgamation of clinical, laboratory, and sociodemographic data. The Random Forest (RF) algorithm emerged as the best performer, achieving a robust classification accuracy of 89% and an AUC of 0.87. This study solidifies the foundational premise that machine learning can effectively discern depression from a combination of diverse data types.

Next, Liu et al. [93] introduced EarlyDetect (ED), a composite screening application utilizing machine learning to incorporate a wide spectrum of variables. From family history of mental illness to suicide ideation, ED exemplifies a comprehensive approach. Using the ElasticNet algorithm, it achieved a balanced accuracy of 72% with an AUC of 0.781. This underlines the potential for machine learning to be effective in complex, real-world settings, synthesizing multiple factors into predictive models.

Adding to the narrative, Ma et al. [94] strived to create a machine learning framework that could expedite the Affective Disorder Evaluation scale. The resulting Bipolar Diagnosis Checklist in Chinese (BDCC), which used the RF algorithm to rank feature importance, achieved an outstanding classification accuracy of 99.6%. The success of BDCC emphasises the role machine learning can play in simplifying and accelerating mental health evaluations.

Simultaneously, Mato-Abad et al. [95] leveraged Artificial Neural Networks (ANNs) to identify a subtype of mild cognitive impairment (MCI) associated with depression. The ANN's success, with an 86% accuracy, highlights the potential of machine learning in elucidating the nuanced intersections between different mental health conditions.

On a parallel track, Meng et al. [96] introduced a temporal deep learning model performing bi-directional representation learning on Electronic Health Record (EHR) sequences. The model's AUC ranged from 0.73 - 0.85, based on the prediction window timeframe. This exploration of temporal modelling in EHR data showcases machine learning's ability to draw insights from longitudinal health data.

In another initiative, Meng et al. [97] devised a model incorporating temporal Hierarchical Clinical Embeddings with Topic Modelling (HCET), addressing data sparsity issues. The improvement in AUCs further emphasized the potential of machine learning in handling complex, sparse datasets.

Adding another dimension, Parker et al. [98] sought to discriminate between bipolar and unipolar subjects. They achieved a promising classification accuracy of 96%, demonstrating machine learning's capacity to distinguish between different mental health disorders, even within the challenging context of unbalanced datasets.

Sharma and Verbeke [99] employed the Extreme Gradient Boosting algorithm on a biomarker dataset, achieving a balanced classification accuracy of 94.42% despite initial dataset biases. This finding underscores machine learning's robustness and adaptability in the face of imbalanced data.

Lastly, Zhou et al. [100] leveraged natural language processing in analysing discharge summaries of depressed patients. Their system, MTERMS, consistently outperformed standard classifiers, reaffirming the strength of machine learning in interpreting unstructured text data.

Together, these studies craft a compelling tale of how machine learning has been applied to diverse data types and challenges in depression detection, consistently achieving impressive results. It illustrates a trend towards increasingly complex and real-world applicable models, with promising indications for the future of machine learning in mental health diagnostics.

B. WEARABLE TECHNIQUES

A vast amount of research has been conducted into detecting depression using a wearable sensors that produce EEG signals. Table 7 is summarising the best performing studies with the most popular classification method being CNNs and SVMs.

Following on, different classification methods have been used with EEG signals. Firstly, Cai et al. [101] introduced a multimodal model fusing different EEG data sources, gathered under a range of emotional conditions. Feature weighting was performed through a genetic algorithm on linear and non-linear features, with this unique approach leading to a robust classification accuracy of 86.98%. This method emphasizes the potential of multimodal models in detecting depression and accentuates the strength of genetic algorithms in feature weighting.

Akbari et al. used k-Nearest Neighbours (k-NN) with geometric features extracted from the EEG signals' Self-Organising Decision Process (SODP) [102]. The Binary Particle Swarm Optimisation (BPSO) algorithm was utilised for feature selection, culminating in impressive results: 98.79% classification accuracy, 97.72% sensitivity, and 99.86% specificity. This work underscores the value of geometric features and the effectiveness of the BPSO algorithm in feature selection for depression detection.

Moreover, Li et al. [103] ventured to extract multiple linear and non-linear features from EEG signals. A rigorous comparison of five different feature selection methods was carried out, with significant discriminant features being identified using Bonferroni correction t-tests. The outcome was a commendable average classification accuracy of 95%, highlighting the importance of meticulous feature selection in achieving high classification accuracy.

Simultaneously, Saeedi et al. applied sample and approximate entropy to wavelet packets, with significant features selected using a Genetic Algorithm (GA) [104]. This method achieved a classification accuracy, sensitivity, and specificity of 98.44%, 97.10%, and 100% respectively. The use of GA once again demonstrates its potency in feature selection, enhancing the classification performance.

Furthermore, research has also delved into the utilization of MLPs in classifying EEG data. Ahmadlou et al. employed a wavelet-chaos methodology, using Katz's and Higuchi's fractal dimensions as measures of nonlinearity and complexity [105]. The resulting 91.3% classification accuracy

TABLE 7. The list of studies where EEG signals have been analysed to detect depression. '-' denotes that the author has not specified a value and '*' denotes an average value. Different datasets have been used throughout the studies, but the table concisely summarises the effectiveness of EEG signals for classification of Depression.

Classification Technique	Methodology	Accuracy (%)	Sensitivity (%)	Specificity (%)	Cross-Validation
CNN	The CNN model was made up of 5 convolution layers, 5 pooling layers and 3 fully-connected layers with a filter size of 5 [134].	94.51*	93.44	95.59	✓
	A deep hybrid model using a CNN to learn the temporal properties of the signal which is then passed to a LSTM to learn the sequences [45].	98.39*	99.13	99.13	✓
	EEG signals are used to approximate the effective connectivity in the brain default mode network (DMN). The effective connections between the six major regions within the DMN are inputted to a 3D-CNN [135].	100	100	100	✓
	Investigated a variety of different ConvNet architectures with different inputs. One input was a trial-wise strategy where the entire signal is inputted. The other is a frame-wise strategy, where the signal is cropped into 1-second length frames [136].	98.87	-	-	✓
	Generating connectivity features using frequency-domain extended multivariate autoregression as an input to a pretrained ResNet-50. An LSTM is then exploited to capture depression specific information [44].	90.22	90.31	90.14	✓
	Modelled 3 different CNN architectures on the 4 main frequency bands of an EEG signal. Further features such as spatial resolution with location information were analysed [137].	91*	-	-	✓
	Power spectral density and activity were respectively extracted as features using an Auto-regress model and Hjorthalgorithm. Spatial information was added of the node placements before being processed by a CNN [138].	84.75	-	-	-
	A CNN is used to generate feature maps of the EEG signal. The feature maps are then inputted into an LSTM to learn different patterns in the signal [139].	98.96*	99.06	98.83	✓
	A Short-Time Fourier transform is applied to the EEG Signal to create spectrogram images. The images are used as an input to the CNN [140].	99.58	99.70	99.48	✓
	Segmenting the EEG signal in one second window size bits allowed for the input to the CNN. Using a deep CNN with several layers the signal is processed before a decision is made [141].	98.32	-	-	✓
	Effective brain connectivity shows relationships in EEG which are extracted by Generalised Partial Directed Coherence and direct Directed transfer function methods. The featuresare then inputted into different model architectures [142].	99.25	98.52	100	✓
	Created a CNN model called DeprNet which has a normalised image of the raw EEG data as its input. A 1D convolution is applied on the time dimension which allows the spatial dimension to be untouched [143].	99.37	88.70	-	✓

TABLE 7. *(Continued.)* **The list of studies where EEG signals have been analysed to detect depression. '-' denotes that the author has not specified a value and '*' denotes an average value. Different datasets have been used throughout the studies, but the table concisely summarises the effectiveness of EEG signals for classification of Depression.**

	Created a model named DepHNN (Depression Hybrid Neural Network). It exploits a CNN for temporal learning of the EEG signal which is then fed into a LSTM for sequence learning [144].	99.10	-	-	✓
SVM	A novel Depression Diagnosis Index is introduced through a combination of 7 nonlinear features. The features are further ranked using t-tests. A polynomial kernel was used. [145].	98	97	98.5	-
	Derived geometric features from the EEG signals second-order differential plot (SODP). The Binary Particle Swarm Optimization (BPSO) algorithm was utilised to choose the suitable features. A Radial Basis Function (RBF) was introduced [102].	97.63*	96.81	98.44	✓
	Proposed a method based on empirical wavelet transforms (EWT) and centered correntropy (CC). The EWT extracts the rhythms in the signal while the CC is computed to disriminate features. [146].	98.76	98.47	99.05	✓
	Generated multi-leveled features using discrete wavelet transforms (DWT) and Melamine patterns. Using the neighbourhood component analysis algorithm, the most relevant features can be selected as an input. [147].	99.05	99.27	99.87	✓
	DWT is performed up to two levels to obtain features extracted from various coefficient levels of the DWT. The Student's t-test is then performed to understand what are the significant differences in the features [148].	88.92	87.88	90.09	✓
	Local Binary Pattern (LBP) analysis was conducted by encoding the segmented signal. Signal Spectrum Analysis was then performed to decompose and reconstruct the LBP signal. This removes noise and divides the frequency band [149].	99.24	99.34	99.12	✓
	Power spectral density and activity were respectively extracted as features using an Auto-regress model and Hjorth algorithm. A deep forest transformed the original features to potentially improve feature engineering [138].	89.02*	-	-	-
	Utilised the phase lag index (PLI) method to calculate connectivity matrices. Graph theory-basedmethods were then exploited to measure the topology of brain networks across the major frequency bands [150].	89.70	89.40	89.90	✓
	Linear and nonlinear features were extracted from the temporal region of the brain using 6 channels. The ReliefF algorithm is used to judge the quality of a feature based on its discriminatory power with samples close by [151].	96.02	-	-	-
	Feature extraction is accomplished by applying a continous wavelet transform (CWT) on each signal.The features are then dimensionally reduced using both principle and kernel principle component analysis while being ranked using Student's t-test [152].	99.26	99.10	99.43	✓
	Exploited synchronization likelihood (SL) for extracting features from the EEG signal. The SL was calculated for each channel pair in the frontal, temporal, occipital and central lobes. Feature selection involving a rank based method based onthe ROC was implemented [153].	98.0	99.9	99.50	✓

illuminates the potential of fractal dimensions and chaos theory in depression detection.

In a parallel effort, Cukic et al. examined Higuchi's Fractal Dimension and Sample Entropy as non-linear measures in discriminating between depressed patients and controls [106]. By leveraging Principal Component Analysis (PCA) for feature dimensionality reduction, they achieved an average classification accuracy of 97.56%. This reinforces the idea that non-linear measures can be highly discriminative and the role of dimensionality reduction techniques in boosting classification performance.

Overall, these studies show the exploration into the detection of depression using EEG signals. They illustrate the evolution of methodologies, from the use of different feature extraction techniques to the application of various machine learning algorithms. The consistently high classification accuracies across studies reinforce the potential of these approaches in advancing depression detection.

VII. CONCLUSION AND FUTURE WORK

This survey has gone into detail about machine learning applications in mental health detection. It can be observed that the most popular methods for automatic detection of depression and ADHD is by exploiting imaging data and EEG data. The non-intrusive nature of the EEG provides an argument that it is the preferred choice. This is due to the vast amount of methods that can be applied to analysis, while causing no harm to the subject.

The biggest drawback about research involving mental health conditions is the size of the dataset. Due to the nature of the conditions, for both ADHD and depression it is difficult to get enough subjects to participate in the research. Furthermore there are possible implications with protecting the privacy of all subjects due to it being very sensitive data. When subjects have agreed to have their data used, there is also the issue of whether the data can be publicly shared or whether it remains private. Lastly, with regards to ADHD and depression, the spectrum of behaviour is vast, meaning some behaviour is very rigid or too excitatory. Therefore, training a classifier to detect these behaviours can be even harder as there is not enough data to cover such a vast spectrum.

Following on, there is more research being conducted into depression. This could be due to the awareness of the mental health condition being bigger or because of the available datasets. We suggest that for both ADHD and Depression respectively, there is a collective movement for a joint database containing multimodal data for the respective mental health conditions. Within these databases, there would be an established method for protecting the participants privacy such as converting their identity to a number/letter and processing the video/image data using techniques such as the Histogram of Gradients. The file types would be made consistent so that all users would know what to expect and baseline scores would be achieved to provide state-of-the-art comparisons. Lastly, for use in research, an End User Licence (EULA) would have to be signed to protect the organisers and subjects' data that is involved within the dataset.

Machine learning is transforming the landscape of ADHD and depression detection and classification through innovative data collection and analysis methods. These encompass imaging techniques, processing of medical notes, and wearable technology, reflecting ADHD's complex nature and showcasing machine learnings potential in diagnosis and treatment.

ADHD diagnosis has seen successful employment of imaging techniques, leveraging SVM and deep learning models. Despite needing large data sets and often dealing with unbalanced ADHD-200 datasets, these challenges are overcome using data augmentation and hypothesis testing frameworks. High classification accuracies from multiple studies reinforce the value of imaging data in ADHD detection. ML has also proven successful in extracting rich clinical information from medical notes, with Decision Trees, SVMs, and hybrid AI models delivering impressive classification accuracy. While there are issues like overfitting and data heterogeneity, these applications highlight the role of AI in clinical decision-making. Incorporating wearable technology provides a non-invasive means of collecting EEG signals for ADHD classification. Techniques such as CNNs and SVMs have been effective in analyzing this data. However, ensuring the models' applicability to new patients and real-world conditions remains a challenge.

For depression detection, machine learning has similarly demonstrated remarkable adaptability and effectiveness. Brain imaging data, clinical notes, sociodemographic data, laboratory data, wearable sensor data, and electronic health records have all been effectively utilized. Algorithms such as SVMs, Random Forest, ElasticNet, Extreme Gradient Boosting, and Artificial Neural Networks have yielded high accuracy rates across diverse data sources. Moreover, machine learning's success in discerning between different depressive disorders could revolutionize personalized treatment.

However, the quality of machine learning models is contingent on the quality of data they're trained on. Continued efforts are essential to ensure the robustness and applicability of these models across various populations and settings. The Intelligent Sensing Group at Newcastle University is conducting their own Intelligent Sensing ADHD trial (ISAT) that involves audio-visual data of controls and ADHD subjects. The aim is for this data to be publicly available once correctly processed.

In conclusion, machine learning offers substantial potential for improving ADHD and depression diagnostics. Despite challenges related to data quality, overfitting, and algorithm interpretability, machine learnings ability to identify patterns in complex datasets makes it a valuable tool in mental health research. Future efforts should focus on creating reliable models, protecting patient data, and ensuring models can be generalized to different populations. Effective collaboration between clinicians, data scientists, and patients will be key to maximizing machine learning's potential in mental health diagnosis and treatment.

REFERENCES

[1] J. W. Kanter, A. M. Busch, C. E. Weeks, and S. J. Landes, ''The nature of clinical depression: Symptoms, syndromes, and behavior analysis,'' *Behav. Anal.*, vol. 31, no. 1, pp. 1–21, Apr. 2008.

[2] K. Low. *Understanding Hypersensitivity in ADHD*. Accessed: Jan. 6, 2023. [Online]. Available: https://www.verywellmind.com/sensitivities-and-adhd-20473

[3] World Health Organisation. *Depressive Disorder (Depression)*. Accessed: Jan. 6, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[4] R. Thomas, S. Sanders, J. Doust, E. Beller, and P. Glasziou, ''Prevalence of attention-deficit/hyperactivity disorder: A systematic review and meta-analysis,'' *Pediatrics*, vol. 135, no. 4, pp. e994–e1001, Apr. 2015.

[5] P. Song, M. Zha, Q. Yang, Y. Zhang, X. Li, and I. Rudan, ''The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis,'' *J. Global Health*, vol. 11, Feb. 2021, Art. no. 04009.

[6] U. P. Ramtekkar, A. M. Reiersen, A. A. Todorov, and R. D. Todd, ''Sex and age differences in attention-deficit/hyperactivity disorder symptoms and diagnoses: Implications for DSM-V and ICD-11,'' *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 49, no. 3, pp. 217–283, 2010.

[7] Z. Huntley, S. Maltezos, C. Williams, A. Morinan, A. Hammon, D. Ball, E. J. Marshall, F. Keaney, S. Young, P. Bolton, K. Glaser, R. Howe-Forbes, J. Kuntsi, K. Xenitidis, D. Murphy, and P. J Asherson, ''Rates of undiagnosed attention deficit hyperactivity disorder in London drug and alcohol detoxification units,'' *BMC Psychiatry*, vol. 12, p. 223, Dec. 2012.

[8] BBC. *ADHD Diagnosis for Adults 'Can Take Seven Years*. Accessed: Jan. 3, 2022. [Online]. Available: https://www.bbc.co.uk/news/uk-england-44956540#:~:text=About%201.5%20million%20adults%20in,only%20120%2C000%20are%20formally%20diagnosed

[9] B. T. Wymbs, W. E. Pelham, B. S. G. Molina, E. M. Gnagy, T. K. Wilson, and J. B. Greenhouse, ''Rate and predictors of divorce among parents of youth with ADHD,'' *J. Consulting Clin. Psychol.*, vol. 76, no. 5, pp. 735–744, Oct. 2008.

[10] K. Doheny, ''Divorce more likely in ADHD families?'' WebMD, Tech. Rep., 2008. [Online]. Available: https://www.cbsnews.com/news/divorce-more-likely-in-adhd-families/

[11] J. A. Doshi, P. Hodgkins, J. Kahle, V. Sikirica, M. J. Cangelosi, J. Setyawan, M. H. Erder, and P. J. Neumann, ''Economic impact of childhood and adult attention-deficit/hyperactivity disorder in the United States,'' *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 51, no. 10, pp. 990–1002, Oct. 2012.

[12] C. Mohr-Jensen and H.-C. Steinhausen, ''A meta-analysis and systematic review of the risks associated with childhood attention-deficit hyperactivity disorder on long-term outcome of arrests, convictions, and incarcerations,'' *Clin. Psychol. Rev.*, vol. 48, pp. 32–42, Aug. 2016.

[13] *Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR)*, 5th ed., Amer. Psychiatric Assoc., Arlington, VA, USA, 2013.

[14] C. K. Conners, J. Pitkanen, and S. R. Rzepa, *Conners (Conners 3; Conners 2008)*, 3rd ed. New York, NY, USA: Springer, 2011, pp. 675–678.

[15] I. Macrory, *Measuring National Well-Being: Life in the UK*. South Wales, Wales: Office for UK National Statistics, 2016.

[16] W. H. Organisation, ''Depression and other common mental disorders: Global health estimates,'' World Health Organisation, Geneva, Switzerland, Tech. Rep. WHO/MSD/MER/2017.2, 2017. [Online]. Available: https://apps.who.int/iris/handle/10665/254610

[17] J. Alonso et al., ''Treatment gap for anxiety disorders is global: Results of the world mental health surveys in 21 countries,'' *Depression Anxiety*, vol. 35, no. 3, pp. 195–208, Mar. 2018.

[18] T. Beiwinkel, S. Kindermann, A. Maier, C. Kerl, J. Moock, G. Barbian, and W. Rössler, ''Using smartphones to monitor bipolar disorder symptoms: A pilot study,'' *JMIR Mental Health*, vol. 3, no. 1, p. e2, Jan. 2016.

[19] U.K.-Government. *Labour Force Survey: Self Reported Work-Related Ill Health Workplace Injuries*. Accessed: Oct. 5, 2021. [Online]. Available: https://www.hse.gov.uk/statistics/lfs/index.html

[20] D. Stevenson and P. Farmer, ''Thriving at work: A review of mental health and employers,'' Dept. Health Social Care UK, London, U.K., Tech. Rep., Oct. 2017. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/658145/thriving-at-work-stevenson-farmer-review.pdf

[21] J. Banks and X. Xu, *The Mental Health Effects of the First Two Months of Lockdown and Social Distancing During the COVID-19 Pandemic in the UK*. Linköping, Sweden: IFS, Jun. 2020.

[22] S. Gallagher and M. A. Wetherell, ''Risk of depression in family caregivers: Unintended consequence of COVID-19,'' *BJPsych Open*, vol. 6, no. 6, pp. 1–5, 2020, Art. no. e119.

[23] R. Stewart, A. Jewell, M. Broadbent, I. Bakolis, and J. Das-Munshi, ''Causes of death in mental health service users during the first wave of the COVID-19 pandemic: South London and Maudsley data from March to June 2020, compared with 2015–2019,'' *MedRxiv*, doi: 10.1101/2020.10.25.20219071.

[24] H. Yang, W. Chen, Y. Hu, Y. Chen, Y. Zeng, Y. Sun, Z. Ying, J. He, Y. Qu, D. Lu, F. Fang, U. Valdimarsdóttir, and H. Song, ''Pre-pandemic psychiatric disorders and risk of COVID-19: A cohort analysis in the U.K. biobank,'' Lancet Healthy Longevity, Tech. Rep., 2020, vol. 1, no. 2. [Online]. Available: https://www.thelancet.com/journals/lanhl/article/PIIS2666-7568(20)30013-1/fulltext

[25] J. S. Evert, C. C. Hannie, J. J. S. Kooij, M. Marieke, T. F. B. Aartjan, and J. H. D. Dorly, ''The role of adverse life events on depression in older adults with ADHD,'' *J. Affect. Disorders*, vol. 174, pp. 574–579, Mar. 2015.

[26] A. J. Lundervold, S. P. Hinshaw, L. Sørensen, and M.-B. Posserud, ''Co-occurring symptoms of attention deficit hyperactivity disorder (ADHD) in a population-based sample of adolescents screened for depression,'' *BMC Psychiatry*, vol. 16, no. 1, pp. 1–10, Dec. 2016.

[27] C. Nash, R. Nair, and S. M. Naqvi, ''Machine learning and ADHD mental health detection—A short survey,'' in *Proc. 25th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2022, pp. 1–8.

[28] D. S. B. A. Hamid, S. B. Goyal, and P. Bedi, ''Integration of deep learning for improved diagnosis of depression using EEG and facial features,'' *Mater. Today, Proc.*, vol. 80, pp. 1965–1969, Jan. 2023.

[29] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, ''Machine learning in mental health: A scoping review of methods and applications,'' *Psychol. Med.*, vol. 49, no. 9, pp. 1426–1448, Jul. 2019.

[30] S. Graham, C. Depp, E. E. Lee, C. Nebeker, X. Tu, H.-C. Kim, and D. V. Jeste, ''Artificial intelligence for mental health and mental illnesses: An overview,'' *Current Psychiatry Rep.*, vol. 21, no. 11, p. 116, Nov. 2019.

[31] H. W. Loh, C. P. Ooi, P. D. Barua, E. E. Palmer, F. Molinari, and U. R. Acharya, ''Automated detection of ADHD: Current trends and future perspective,'' *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105525.

[32] N. Sethu and R. Vyas, ''Overview of machine learning methods in ADHD prediction,'' in *Advances in Bioengineering*. Springer, May 2020, pp. 51–71.

[33] *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*, World Health Organization, Geneva, Switzerland, 1993.

[34] *Attention Deficit Hyperactivity Disorder*, National Institute of Mental Health, Bethesda, MD, USA, 2022.

[35] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, ''The PHQ-9: Validity of a brief depression severity measure,'' *J. Gen. Internal Med.*, vol. 16, no. 9, pp. 606–613, Sep. 2001.

[36] J. Ware and C. D. Sherbourne, ''The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection,'' *Med. Care*, vol. 30, no. 6, pp. 473–483, 1992.

[37] J. Upton, *Beck Depression Inventory (BDI)*. New York, NY, USA: Springer, 2013, pp. 178–179.

[38] G. H. Glover, ''Overview of functional magnetic resonance imaging,'' *Neurosurg. Clinics North Amer.*, vol. 22, no. 2, p. 133–139, 2011.

[39] M. R. G. Brown, G. S. Sidhu, R. Greiner, N. Asgarian, M. Bastani, P. H. Silverstone, A. J. Greenshaw, and S. M. Dursun, ''ADHD-200 global competition: Diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements,'' *Frontiers Syst. Neurosci.*, vol. 6, p. 69, Sep. 2012.

[40] N. Cauli and D. Reforgiato Recupero, ''Survey on videos data augmentation for deep learning models,'' *Future Internet*, vol. 14, no. 3, p. 93, Mar. 2022.

[41] *AVEC 19: Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, Association for Computing Machinery, New York, NY, USA, 2019.

[42] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, "AVEC 2013: The continuous audio/visual emotion and depression recognition challenge," in *Proc. 3rd ACM Int. Workshop Audio/Vis. Emotion Challenge*, Barcelona, Spain, 2013. [Online]. Available: https://www.researchgate.net/publication/262157517_AVEC_2013_-_The_continuous_AudioVisual_Emotion_and_depression_recognition_challenge

[43] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic, "AVEC 2014: 3D dimensional affect and depression recognition challenge," in *Proc. 4th Int. Workshop Audio/Visual Emotion Challenge (AVEC)*, Orlando, FL, USA, 2014. [Online]. Available: https://www.researchgate.net/publication/287702683_AVEC_2014_-_3D_dimensional_affect_and_depression_recognition_challenge#fullTextFileContent

[44] C. Uyulan, S. de la Salle, T. T. Erguzel, E. Lynn, P. Blier, V. Knott, M. M. Adamson, M. Zelka, and N. Tarhan, "Depression diagnosis modeling with advanced computational methods: Frequency-domain eMVAR and deep learning," *Clin. EEG Neurosci.*, vol. 53, no. 1, pp. 24–36, Jan. 2022.

[45] B. Ay, O. Yildirim, M. Talo, U. B. Baloglu, G. Aydin, S. D. Puthankattil, and U. R. Acharya, "Automated depression detection using deep representation and sequence learning with EEG signals," *J. Med. Syst.*, vol. 43, no. 7, p. 205, Jul. 2019.

[46] Y. Tang, X. Li, Y. Chen, Y. Zhong, A. Jiang, and C. Wang, "High-accuracy classification of attention deficit hyperactivity disorder with $l_{2,1}$-norm linear discriminant analysis and binary hypothesis testing," *IEEE Access*, vol. 8, pp. 56228–56237, 2020.

[47] S. Sartipi, H. Kalbkhani, P. Ghasemzadeh, and M. G. Shayesteh, "Stockwell transform of time-series of fMRI data for diagnoses of attention-deficit hyperactivity disorder," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105905.

[48] J. C. Bledsoe, C. Xiao, A. Chaovalitwongse, S. Mehta, T. J. Grabowski, M. Semrud-Clikeman, S. Pliszka, and D. Breiger, "Diagnostic classification of ADHD versus control: Support vector machine classification using brief neuropsychological assessment," *J. Attention Disorders*, vol. 24, no. 11, pp. 1547–1556, 2020.

[49] T. Chen, G. Antoniou, M. Adamou, I. Tachmazidis, and P. Su, "Automatic diagnosis of attention deficit hyperactivity disorder using machine learning," *Appl. Artif. Intell.*, vol. 35, no. 9, pp. 657–669, Jul. 2021.

[50] I. Tachmazidis, T. H. Chen, M. Adamou, and G. Antoniou, "A hybrid AI approach for supporting clinical diagnosis of attention-deficit hyperactivity disorder (ADHD) in adults," *Health Inf. Sci. Syst.*, vol. 9, no. 1, pp. 1–8, 2021.

[51] H. Christiansen, M.-L. Chavanon, O. Hirsch, M. H. Schmidt, C. Meyer, A. Müller, H.-J. Rumpf, I. Grigorev, and A. Hoffmann, "Use of machine learning to classify adult ADHD and other conditions based on the Conners' adult ADHD rating scales," *Sci. Rep.*, vol. 10, Nov. 2020, Art. no. 18871.

[52] M. Duda, N. Haber, J. Daniels, and D. P. Wall, "Crowdsourced validation of a machine-learning classification system for autism and ADHD," *Transl. Psychiatry*, vol. 7, May 2017, Art. no. e1133.

[53] P. Amado-Caballero, P. Casaseca-de-la-Higuera, S. Alberola-Lopez, J. M. Andres-de-Llano, J. A. L. Villalobos, J. R. Garmendia-Leiza, and C. Alberola-Lopez, "Objective ADHD diagnosis using convolutional neural networks over daily-life activity records," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2690–2700, Sep. 2020.

[54] H. Chen, W. Chen, Y. Song, L. Sun, and X. Li, "EEG characteristics of children with attention-deficit hyperactivity disorder," *Neuroscience*, vol. 406, pp. 444–456, May 2019.

[55] M. Moghaddari, M. Z. Lighvan, and S. Danishvar, "Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG," *Comput. Methods Programs Biomed.*, vol. 197, Dec. 2020, Art. no. 105738.

[56] A. Ahmadi, M. Kashefi, H. Shahrokhi, and M. A. Nazari, "Computer aided diagnosis system using deep convolutional neural networks for ADHD subtypes," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102227.

[57] M.-Y. Chang, C.-S. Ouyang, C.-T. Chiang, R.-C. Yang, R.-C. Wu, H.-C. Wu, and L.-C. Lin, "A new method of diagnosing attention-deficit hyperactivity disorder in male patients by quantitative EEG analysis," *Clin. EEG Neurosci.*, vol. 50, no. 5, pp. 339–347, Sep. 2019.

[58] H. Chen, J. Yan, Y. Gu, Y. Song, and X. Li, "Mutual information analysis of EEG of children with attention-deficit hyperactivity disorder," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 2342–2347.

[59] F. D. Dea, M. Ajcevic, M. Stecca, C. Zanus, M. Carrozzi, A. Cuzzocrea, and A. Accardo, "A big-data-analytics framework for supporting classification of ADHD and healthy children via principal component analysis of EEG sleep spindles power spectra," *Proc. Comput. Sci.*, vol. 159, pp. 1584–1590, 2019. [Online]. Available: https://www.researchgate.net/publication/336537750_A_Big-Data-Analytics_Framework_for_Supporting_Classification_of_ADHD_and_Healthy_Children_via_Principal_Component_Analysis_of_EEG_Sleep_Spindles_Power_Spectra, doi: 10.1016/j.procs.2019.09.329.

[60] M. Rezaeezadeh, S. Shamekhi, and M. Shamsi, "Attention deficit hyperactivity disorder diagnosis using non-linear univariate and multivariate EEG measurements: A preliminary study," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 577–592, Jun. 2020.

[61] H. T. Tor, C. P. Ooi, N. S. Lim-Ashworth, J. K. E. Wei, V. Jahmunah, S. L. Oh, U. R. Acharya, and D. S. S. Fung, "Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105941.

[62] A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Müller, and G. Candrian, "Machine learning approach for classification of ADHD adults," *Int. J. Psychophysiol.*, vol. 93, no. 1, pp. 162–166, Jul. 2014.

[63] S. S. Poil, S. Bollmann, C. Ghisleni, R. L. O'Gorman, P. Klaver, J. Ball, D. Eich-Höchli, D. Brandeis, and L. Michels, "Age dependent electroencephalographic changes in attention-deficit hyperactivity disorder (ADHD)," *Clin Neurophysiol*, vol. 125, no. 8, pp. 1626–1638, 2014.

[64] Y. Zhu, Y. Shang, Z. Shao, and G. Guo, "Automated depression diagnosis based on deep networks to encode facial appearance and dynamics," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 578–584, Oct. 2018.

[65] P. Meshram and R. K. Rambola, "Diagnosis of depression level using multimodal approaches using deep learning techniques with multiple selective features," *Expert Syst.*, vol. 40, no. 4, May 2023.

[66] L. He, D. Jiang, and H. Sahli, "Automatic depression analysis using dynamic facial appearance descriptor and Dirichlet process Fisher encoding," *IEEE Trans. Multimedia*, vol. 21, no. 6, pp. 1476–1486, Jun. 2019.

[67] X. Li, T. Cao, S. Sun, B. Hu, and M. Ratcliffe, "Classification study on eye movement data: Towards a new approach in depression detection," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2016, pp. 1227–1232.

[68] Q.-B. Hong, C.-H. Wu, M.-H. Su, and C.-C. Chang, "Exploring macroscopic and microscopic fluctuations of elicited facial expressions for mood disorder classification," *IEEE Trans. Affect. Comput.*, vol. 12, no. 4, pp. 989–1001, Oct. 2021.

[69] X. Zhou, K. Jin, Y. Shang, and G. Guo, "Visually interpretable representation learning for depression recognition from facial images," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 542–552, Jul. 2020.

[70] M. Tadalagi and A. M. Joshi, "AutoDep: Automatic depression detection using facial expressions based on linear binary pattern descriptor," *Med. Biol. Eng. Comput.*, vol. 59, no. 6, pp. 1339–1354, Jun. 2021.

[71] Y. Shang, Y. Pan, X. Jiang, Z. Shao, G. Guo, T. Liu, and H. Ding, "LQGDNet: A local quaternion and global deep network for facial depression recognition," *IEEE Trans. Affect. Comput.*, early access, Dec. 31, 2021, doi: 10.1109/TAFFC.2021.3139651.

[72] M. A. Uddin, J. B. Joolee, and Y.-K. Lee, "Depression level prediction using deep spatiotemporal features and multilayer bi-LTSM," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 864–870, Apr. 2022.

[73] S. Song, S. Jaiswal, L. Shen, and M. Valstar, "Spectral representation of behaviour primitives for depression analysis," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 829–844, Apr. 2022.

[74] L. Yang, D. Jiang, and H. Sahli, "Integrating deep and shallow models for multi-modal depression analysis—Hybrid architectures," *IEEE Trans. Affect. Comput.*, vol. 12, no. 1, pp. 239–253, Jan. 2021.

[75] W. C. de Melo, E. Granger, and M. B. López, "MDN: A deep maximization-differentiation network for spatio-temporal depression detection," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 578–590, Jan. 2023.

[76] A. C. Arevian, D. Bone, N. Malandrakis, V. R. Martinez, K. B. Wells, D. J. Miklowitz, and S. Narayanan, "Clinical state tracking in serious mental illness through computational analysis of speech," *PLoS ONE*, vol. 15, no. 1, Jan. 2020, Art. no. e0225695.

[77] H. Chen, Y. Lin, Y. Li, W. Wang, P. Wang, and Y. Lei, "Hybrid feature embedded sparse stacked autoencoder and manifold dimensionality reduction ensemble for mental health speech recognition," *IEEE Access*, vol. 9, pp. 28729–28741, 2021.
[78] N. Cummins, V. Sethu, J. Epps, S. Schnieder, and J. Krajewski, "Analysis of acoustic space variability in speech affected by depression," *Speech Commun.*, vol. 75, pp. 27–49, Dec. 2015.
[79] L. He and C. Cao, "Automated depression analysis using convolutional neural networks from speech," *J. Biomed. Informat.*, vol. 83, pp. 103–111, Jul. 2018.
[80] H. Jiang, B. Hu, Z. Liu, G. Wang, L. Zhang, X. Li, and H. Kang, "Detecting depression using an ensemble logistic regression model based on multiple speech features," *Comput. Math. Methods Med.*, vol. 2018, Sep. 2018, Art. no. 6508319.
[81] J. Li, X. Fu, Z. Shao, and Y. Shang, "Improvement on speech depression recognition based on deep networks," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2018, pp. 2705–2709.
[82] M. Muzammel, H. Salam, Y. Hoffmann, M. Chetouani, and A. Othmani, "AudVowelConsNet: A phoneme-level based deep CNN architecture for clinical depression diagnosis," *Mach. Learn. with Appl.*, vol. 2, Dec. 2020, Art. no. 100005.
[83] Z. Zhao, Z. Bao, Z. Zhang, J. Deng, N. Cummins, H. Wang, J. Tao, and B. Schuller, "Automatic assessment of depression from speech via a hierarchical attention transfer network and attention autoencoders," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 2, pp. 423–434, Feb. 2020.
[84] L. Cao, S. Guo, Z. Xue, Y. Hu, H. Liu, T. E. Mwansisya, W. Pu, B. Yang, C. Liu, J. Feng, E. Y. Chen, and Z. Liu, "Aberrant functional connectivity for diagnosis of major depressive disorder: A discriminant analysis," *Psychiatry Clin. Neurosci.*, vol. 68, no. 2, pp. 110–119, 2014.
[85] H. Guo, X. Cao, Z. Liu, H. Li, J. Chen, and K. Zhang, "Machine learning classifier using abnormal brain network topological metrics in major depressive disorder," *Neuroreport*, vol. 23, no. 17, pp. 1006–1011, 2012.
[86] H. Guo, M. Qin, J. Chen, Y. Xu, and J. Xiang, "Machine-learning classifier for patients with major depressive disorder: Multifeature approach based on a high-order minimum spanning tree functional brain network," *Comput. Math. Methods Med.*, vol. 2017, Dec. 2017, Art. no. 4820935.
[87] H. Li, L. Cui, L. Cao, Y. Zhang, Y. Liu, W. Deng, and W. Zhou, "Identification of bipolar disorder using a combination of multimodality magnetic resonance imaging and machine learning techniques," *BMC Psychiatry*, vol. 20, no. 1, pp. 1–12, Dec. 2020.
[88] M. J. Rosa, L. Portugal, T. Hahn, A. J. Fallgatter, M. I. Garrido, J. Shawe-Taylor, and J. Mourao-Miranda, "Sparse network-based models for patient classification using fMRI," *NeuroImage*, vol. 105, pp. 493–506, Jan. 2015.
[89] Y. Li, X. Dai, H. Wu, and L. Wang, "Establishment of effective biomarkers for depression diagnosis with fusion of multiple resting-state connectivity measures," *Frontiers Neurosci.*, vol. 15, Sep. 2021, Art. no. 729958.
[90] B. Sen, K. R. Cullen, and K. K. Parhi, "Classification of adolescent major depressive disorder via static and dynamic connectivity," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2604–2614, Jul. 2021.
[91] R. Wang, Y. Hao, Q. Yu, M. Chen, I. Humar, and G. Fortino, "Depression analysis and recognition based on functional near-infrared spectroscopy," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 12, pp. 4289–4299, Dec. 2021.
[92] E. M. de Souza Filho, H. C. V. Rey, R. M. Frajtag, D. M. A. Cook, L. N. D. de Carvalho, A. L. P. Ribeiro, and J. Amaral, "Can machine learning be useful as a screening tool for depression in primary care?" *J. Psychiatric Res.*, vol. 132, pp. 1–6, Jan. 2021.
[93] Y. Liu, J. Hankey, B. Cao, and P. Chokka, "Screening for major depressive disorder in a tertiary mental health centre using EarlyDetect: A machine learning-based pilot study," *J. Affect. Disorders Rep.*, vol. 3, Jan. 2021, Art. no. 100062.
[94] Y. Ma, J. Ji, Y. Huang, H. Gao, Z. Li, W. Dong, S. Zhou, Y. Zhu, W. Dang, T. Zhou, H. Yu, B. Yu, Y. Long, L. Liu, G. Sachs, and X. Yu, "Implementing machine learning in bipolar diagnosis in China," *Transl. Psychiatry*, vol. 9, no. 1, Nov. 2019.
[95] V. Mato-Abad, I. Jiménez, R. García-Vázquez, J. Aldrey, D. Rivero, P. Cacabelos, J. Andrade-Garda, J. Pías-Peleteiro, and S. Rodríguez-Yáñez, "Using artificial neural networks for identifying patients with mild cognitive impairment associated with depression using neuropsychological test features," *Appl. Sci.*, vol. 8, no. 9, p. 1629, Sep. 2018.
[96] Y. Meng, W. Speier, M. K. Ong, and C. W. Arnold, "Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 8, pp. 3121–3129, Aug. 2021.
[97] Y. Meng, W. Speier, M. Ong, and C. W. Arnold, "HCET: Hierarchical clinical embedding with topic modeling on electronic health records for predicting future depression," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1265–1272, Apr. 2021.
[98] G. Parker, M. J. Spoelma, G. Tavella, M. Alda, D. L. Dunner, C. O'Donovan, J. K. Rybakowski, A. Bayes, V. Sharma, P. Boyce, and V. Manicavasagar, "A new machine learning-derived screening measure for differentiating bipolar from unipolar mood disorders," *J. Affect. Disorders*, vol. 299, pp. 513–516, Feb. 2022.
[99] A. Sharma and W. J. M. I. Verbeke, "Improving diagnosis of depression with XGBOOST machine learning model and a large biomarkers Dutch dataset (n = 11,081)," *Frontiers Big Data*, vol. 3, p. 15, Apr. 2020.
[100] L. Zhou, A. W. Baughman, V. J. Lei, K. H. Lai, A. S. Navathe, F. Chang, M. Sordo, M. Topaz, F. R. Zhong, M. Murrali, S. Navathe, and R. A. Rocha, "Identifying patients with depression using free-text clinical documents," in *Proc. MEDINFO eHealth-Enabled Health* (Studies in Health Technology and Informatics), vol. 216, 2015, pp. 629–633.
[101] H. S. Cai, J. S. Han, Y. F. Chen, X. C. Sha, Z. Y. Wang, B. Hu, J. Yang, L. Feng, Z. J. Ding, Y. Q. Chen, and J. Gutknecht, "A pervasive approach to EEG-based depression detection," *Complexity*, vol. 2018, pp. 1–13, 2018.
[102] H. Akbari, M. T. Sadiq, M. Payan, S. S. Esmaili, H. Baghri, and H. Bagheri, "Depression detection based on geometrical features extracted from SODP shape of EEG signals and binary PSO," *Traitement du Signal*, vol. 38, no. 1, pp. 13–26, Feb. 2021.
[103] X. Li, B. Hu, S. Sun, and H. Cai, "EEG-based mild depressive detection using feature selection methods and classifiers," *Comput. Methods Programs Biomed.*, vol. 136, pp. 151–161, Nov. 2016.
[104] M. Saeedi, A. Saeedi, and A. Maghsoudi, "Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals," *Phys. Eng. Sci. Med.*, vol. 43, no. 3, pp. 1007–1018, Sep. 2020.
[105] M. Ahmadlou, H. Adeli, and A. Adeli, "Fractality analysis of frontal brain in major depressive disorder," *Int. J. Psychophysiol.*, vol. 85, no. 2, pp. 206–211, 2012.
[106] M. Čukić, M. Stokić, S. Simić, and D. Pokrajac, "The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification method," *Cogn. Neurodyn.*, vol. 14, no. 4, pp. 443–455, Aug. 2020.
[107] C. Cognition. *Stop Signal Task*. Accessed: Jan. 3, 2022. [Online]. Available: https://www.cambridgecognition.com/cantab/cognitive-tests/memory/stop-signal-task-sst
[108] S. Healthineers. *1.5T MRI Scanners*. Accessed: Jan. 3, 2022. [Online]. Available: https://www.siemens-healthineers.com/en-uk/magnetic-resonance-imaging/0-35-to-1-5t-mri-scanner
[109] g.NAUTICULUS. *g.NAUTICULUS Research*. Accessed: Mar. 1, 2020. [Online]. Available: https://www.gtec.at/product/gnautilus-research/
[110] N. Bajaj, "Wavelets for EEG analysis," in *Wavelet Theory*. IntechOpen, ch. 5, Feb. 2020, doi:10.5772/intechopen.94398.
[111] S. De Silva, S. U. Dayarathna, G. Ariyarathne, D. Meedeniya, and S. Jayarathna, "FMRI feature extraction model for ADHD classification using convolutional neural network," *Int. J. E-Health Med. Commun.*, vol. 12, no. 1, pp. 81–105, Jan. 2021.
[112] C. Dou, S. Zhang, H. Wang, L. Sun, Y. Huang, and W. Yue, "ADHD fMRI short-time analysis method for edge computing based on multi-instance learning," *J. Syst. Archit.*, vol. 111, Dec. 2020, Art. no. 101834.
[113] Z. Mao, Y. Su, G. Xu, X. Wang, Y. Huang, W. Yue, L. Sun, and N. Xiong, "Spatio-temporal deep learning method for ADHD fMRI classification," *Inf. Sci.*, vol. 499, pp. 1–11, Oct. 2019.
[114] J. Peng, M. Debnath, and A. K. Biswas, "Efficacy of novel summation-based synergetic artificial neural network in ADHD diagnosis," *Mach. Learn. With Appl.*, vol. 6, Dec. 2021, Art. no. 100120.
[115] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, "3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI," *IEEE Access*, vol. 5, pp. 23626–23636, 2017.
[116] A. Riaz, M. Asad, E. Alonso, and G. Slabaugh, "DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI," *J. Neurosci. Methods*, vol. 335, Apr. 2020, Art. no. 108506.
[117] Z. Wang, Y. Sun, Q. Shen, and L. Cao, "Dilated 3D convolutional neural networks for brain MRI data classification," *IEEE Access*, vol. 7, pp. 134388–134398, 2019.
[118] S. Dey, A. R. Rao, and M. Shah, "Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects," *Frontiers Neural Circuits*, vol. 8, p. 64, Jun. 2014.

[119] J. Du, L. Wang, B. Jie, and D. Zhang, "Network-based classification of ADHD patients using discriminative subnetwork selection and graph kernel PCA," *Computerized Med. Imag. Graph.*, vol. 52, pp. 82–88, Sep. 2016.

[120] N. A. Khan, S. A. Waheeb, A. Riaz, and X. Shang, "A novel knowledge distillation-based feature selection for the classification of ADHD," *Biomolecules*, vol. 11, no. 8, p. 1093, Jul. 2021.

[121] B. Miao, L. L. Zhang, J. L. Guan, Q. F. Meng, and Y. L. Zhang, "Classification of ADHD individuals and neurotypicals using reliable RELIEF: A resting-state study," *IEEE Access*, vol. 7, pp. 62163–62171, 2019.

[122] L. Shao, Y. Xu, and D. Fu, "Classification of ADHD with bi-objective optimization," *J. Biomed. Informat.*, vol. 84, pp. 164–170, Aug. 2018.

[123] X. Peng, P. Lin, T. Zhang, and J. Wang, "Extreme learning machine-based classification of ADHD using brain structural MRI data," *PLoS ONE*, vol. 8, no. 11, Nov. 2013, Art. no. e79476.

[124] P. Preetha and R. Mallika, "Normalization and deep learning based attention deficit hyperactivity disorder classification," *J. Intell. Fuzzy Syst.*, vol. 40, no. 4, pp. 7613–7621, Apr. 2021.

[125] S. Liu, L. Zhao, J. Zhao, B. Li, and S.-H. Wang, "Attention deficit/hyperactivity disorder classification based on deep spatio-temporal features of functional magnetic resonance imaging," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103239.

[126] T. Eslami and F. Saeed, "Similarity based classification of ADHD using singular value decomposition," in *Proc. 15th ACM Int. Conf. Comput. Frontiers*, 2018, pp. 19–25.

[127] G. Deshpande, P. Wang, D. Rangaprakash, and B. Wilamowski, "Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data," *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2668–2679, Dec. 2015.

[128] S. Ghiassian, R. Greiner, P. Jin, and M. R. G. Brown, "Using functional or structural magnetic resonance images and personal characteristic data to identify ADHD and autism," *PLoS ONE*, vol. 11, no. 12, Dec. 2016, Art. no. e0166934.

[129] A. Anderson, P. K. Douglas, W. T. Kerr, V. S. Haynes, A. L. Yuille, J. Xie, Y. N. Wu, J. A. Brown, and M. S. Cohen, "Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD," *NeuroImage*, vol. 102, pp. 207–219, Nov. 2014.

[130] P. Wang, X. Zhao, J. Zhong, and Y. Zhou, "Localization and diagnosis of attention-deficit/hyperactivity disorder," *Healthcare*, vol. 9, no. 4, p. 372, Mar. 2021.

[131] A. M. S. Aradhya, V. Subbaraju, S. Sundaram, and N. Sundararajan, "Discriminant spatial filtering method (DSFM) for the identification and analysis of abnormal resting state brain activities," *Expert Syst. Appl.*, vol. 181, Nov. 2021, Art. no. 115074.

[132] S. Farzi, S. Kianian, and I. Rastkhadive, "Diagnosis of attention deficit hyperactivity disorder using deep belief network based on greedy approach," in *Proc. 5th Int. Symp. Comput. Bus. Intell. (ISCBI)*, Aug. 2017, pp. 96–99.

[133] N. Qiang, Q. Dong, H. Liang, B. Ge, S. Zhang, C. Zhang, J. Gao, and Y. Sun, "A novel ADHD classification method based on resting state temporal templates (RSTT) using spatiotemporal attention auto-encoder," *Neural Comput. Appl.*, vol. 34, no. 10, pp. 7815–7833, May 2022.

[134] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated EEG-based screening of depression using deep convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 161, pp. 103–113, Jul. 2018.

[135] D. M. Khan, N. Yahya, N. Kamel, and I. Faye, "Automated diagnosis of major depressive disorder using brain effective connectivity and 3D convolutional neural network," *IEEE Access*, vol. 9, pp. 8835–8846, 2021.

[136] X. Li, R. La, Y. Wang, J. Niu, S. Zeng, S. Sun, and J. Zhu, "EEG-based mild depression recognition using convolutional neural network," *Med. Biol. Eng. Comput.*, vol. 57, no. 6, pp. 1341–1352, Jun. 2019.

[137] C. Uyulan, T. T. Ergüzel, H. Unubol, M. Cebi, G. H. Sayar, M. Nezhad Asad, and N. Tarhan, "Major depressive disorder classification based on different convolutional neural network models: Deep learning approach," *Clin. EEG Neurosci.*, vol. 52, no. 1, pp. 38–51, Jan. 2021.

[138] X. Li, X. Zhang, J. Zhu, W. Mao, S. Sun, Z. Wang, C. Xia, and B. Hu, "Depression recognition using machine learning methods with different feature generation strategies," *Artif. Intell. Med.*, vol. 99, Aug. 2019, Art. no. 101696.

[139] P. P. Thoduparambil, A. Dominic, and S. M. Varghese, "EEG-based deep learning model for the automatic detection of clinical depression," *Phys. Eng. Sci. Med.*, vol. 43, no. 4, pp. 1349–1360, Dec. 2020.

[140] H. W. Loh, C. P. Ooi, E. Aydemir, T. Tuncer, S. Dogan, and U. R. Acharya, "Decision support system for major depression detection using spectrogram and convolution neural network with EEG signals," *Expert Syst.*, vol. 39, no. 3, Mar. 2022, Art. no. e12773.

[141] W. Mumtaz and A. Qayyum, "A deep learning framework for automatic diagnosis of unipolar depression," *Int. J. Med. Informat.*, vol. 132, Dec. 2019, Art. no. 103983.

[142] A. Saeedi, M. Saeedi, A. Maghsoudi, and A. Shalbaf, "Major depressive disorder diagnosis based on effective connectivity in EEG signals: A convolutional neural network and long short-term memory approach," *Cogn. Neurodyn.*, vol. 15, no. 2, pp. 239–252, Apr. 2021.

[143] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma, and O. Krejcar, "DeprNet: A deep convolution neural network framework for detecting depression using EEG," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[144] G. Sharma, A. Parashar, and A. M. Joshi, "DepHNN: A novel hybrid neural network for electroencephalogram (EEG)-based screening of depression," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102393.

[145] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. W. Koh, S. D. Puthankatti, and A. Adeli, "A novel depression diagnosis index using nonlinear features in EEG signals," *Eur. Neurol.*, vol. 74, nos. 1–2, pp. 79–83, 2015.

[146] H. Akbari, M. T. Sadiq, and A. U. Rehman, "Classification of normal and depressed EEG signals based on centered correntropy of rhythms in empirical wavelet transform domain," *Health Inf. Sci. Syst.*, vol. 9, no. 1, Dec. 2021.

[147] E. Aydemir, T. Tuncer, S. Dogan, R. Gururajan, and U. R. Acharya, "Automated major depressive disorder detection using melamine pattern with EEG signals," *Int. J. Speech Technol.*, vol. 51, no. 9, pp. 6449–6466, Sep. 2021.

[148] G. M. Bairy, U. C. Niranjan, and S. D. Puthankattil, "Automated classification of depression EEG signals using wavelet entropies and energies," *J. Mech. Med. Biol.*, vol. 16, no. 3, May 2016, Art. no. 1650035.

[149] L. J. Duan, H. L. Liu, H. F. Duan, Y. H. Qiao, and C. M. Wang, "Classification of depression based on local binary pattern and singular spectrum analysis," in *Algorithms and Architectures for Parallel Processing*. Cham, Switzerland: Springer, 2020.

[150] W. Liu, C. Zhang, X. Wang, J. Xu, Y. Chang, T. Ristaniemi, and F. Cong, "Functional connectivity of major depression disorder using ongoing EEG during music perception," *Clin. Neurophysiol.*, vol. 131, no. 10, pp. 2413–2422, Oct. 2020.

[151] S. Mahato, N. Goyal, D. Ram, and S. Paul, "Detection of depression and scaling of severity using six channel EEG data," *J. Med. Syst.*, vol. 44, no. 7, p. 118, Jul. 2020.

[152] U. Raghavendra, A. Gudigar, Y. Chakole, P. Kasula, D. P. Subha, N. A. Kadri, E. J. Ciaccio, and U. R. Acharya, "Automated detection and screening of depression using continuous wavelet transform with electroencephalogram signals," *Expert Syst.*, vol. 40, no. 4, May 2023, Art. no. e12803.

[153] W. Mumtaz, S. S. A. Ali, M. A. M. Yasin, and A. S. Malik, "A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD)," *Med. Biol. Eng. Comput.*, vol. 56, no. 2, pp. 233–246, Feb. 2018.

CHRISTIAN NASH (Student Member, IEEE) received the B.Eng. degree in electrical and computer engineering from Newcastle University, U.K., in 2020, where he is currently pursuing the Ph.D. degree with the Intelligent Sensing and Communication Group.

His research interests include deep learning, medical video analysis, and mental health detection.

RAJESH NAIR is currently a FRCPsych and a Consultant Psychiatrist, leads a team with the Adult ADHD Services, Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust, England. He is also the U.K. Adult ADHD Network (UKAAN) Lead of the North-East of England. He has experience in both National Institute for Health Research (NIHR) and industry funded research projects/trials e.g., digital medicines trial, ketamine in electroconvulsive therapy (ECT) trial, the lithium versus quetiapine in depression (LQD) trial with TEWV, and U.K. landscape study. He has significant expertise with Adult ADHD.

SYED MOHSEN NAQVI (Senior Member, IEEE) received the Ph.D. degree from Loughborough University, U.K., in 2010.

He is currently a Reader (Associate Professor) in multimodal signal and information processing, the Director of the Intelligent Sensing Laboratory, and the Deputy Head of the Intelligent Sensing and Communications Research Group, Newcastle University, Newcastle, U.K. He has above 130 publications in peer-reviewed articles in high impact journals and proceedings of leading international conferences. He organized special sessions in FUSION (2013–2022), delivered seminars, and was a Speaker with University Defence Research Collaboration (UDRC) Summer Schools (2015–2017). He was involved in more than 15 research projects funded by UKRI and Industry (e.g., EPSRC, BBSRC, MoD, UDRC, Thales, Innovate U.K., and NHS). He also successfully supervised and graduated above 20 Ph.D. His research interests include human action, activity, behavior analyses, multiple human target detection, localization, tracking, human speech enhancement and separation, and explainable AI, all for defence and healthcare applications.

Dr. Naqvi is a fellow of the Higher Education Academy. He was an Associate Editor of *Journal of Signal Processing* (Elsevier) (2018–2022). He served two terms of an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING (2019–2023). He has been an Associate Editor of IEEE/ACM TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING, since 2019.