sensors

Systematic Review

Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches

Lin Sze Khoo [1,*], Mei Kuan Lim [2], Chun Yong Chong [2] and Roisin McNaney [1]

[1] Department of Human-Centered Computing, Faculty of Information Technology, Monash University, Clayton, VIC 3800, Australia; roisin.mcnaney@monash.edu
[2] School of Information Technology, Monash University Malaysia, Subang Jaya 46150, Malaysia; lim.meikuan@monash.edu (M.K.L.); chong.chunyong@monash.edu (C.Y.C.)
[*] Correspondence: lin.khoo@monash.edu

Abstract: As mental health (MH) disorders become increasingly prevalent, their multifaceted symptoms and comorbidities with other conditions introduce complexity to diagnosis, posing a risk of underdiagnosis. While machine learning (ML) has been explored to mitigate these challenges, we hypothesized that multiple data modalities support more comprehensive detection and that non-intrusive collection approaches better capture natural behaviors. To understand the current trends, we systematically reviewed 184 studies to assess feature extraction, feature fusion, and ML methodologies applied to detect MH disorders from passively sensed multimodal data, including audio and video recordings, social media, smartphones, and wearable devices. Our findings revealed varying correlations of modality-specific features in individualized contexts, potentially influenced by demographics and personalities. We also observed the growing adoption of neural network architectures for model-level fusion and as ML algorithms, which have demonstrated promising efficacy in handling high-dimensional features while modeling within and cross-modality relationships. This work provides future researchers with a clear taxonomy of methodological approaches to multimodal detection of MH disorders to inspire future methodological advancements. The comprehensive analysis also guides and supports future researchers in making informed decisions to select an optimal data source that aligns with specific use cases based on the MH disorder of interest.

Keywords: machine learning; mental health; multimodal detection; passive sensing; systematic review

check for updates

Citation: Khoo, L.S.; Lim, M.K.; Chong, C.Y.; McNaney, R. Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches. *Sensors* **2024**, *24*, 348. https://doi.org/10.3390/s24020348

Academic Editors: Xin Chen and Andrew King

Received: 25 October 2023
Revised: 14 December 2023
Accepted: 18 December 2023
Published: 6 January 2024

1. Introduction

Mental health (MH) issues are pervasive in modern society, with the World Health Organization estimating that around 1 in 8, or 970 million people, were living with a mental health condition in 2019 [1]. The COVID-19 pandemic brought unprecedented times, leading to a reported increase in rates of anxiety and major depression by 25% in 2020 [2]. Subsequently, 42.9% of people in Australia aged between 16 and 85 years had experienced a mental disorder at some time in their lives as of 2022 [3], whereas 22.8% of adults in the U.S. were estimated to be experiencing mental illness as of 2021 [4]. With figures estimating that MH disorders will contribute to an economic loss of around USD 16 trillion globally by 2030 [5], it is unsurprising that MH has become a government priority worldwide. Specifically, the Comprehensive Mental Health Action Plan 2013–2030 [6] encompasses several global targets to promote improved mental health and well-being, where service coverage for MH conditions will have increased at least by half, and 80% of countries will have integrated mental health into primary health care by 2030. The impacts of MH issues on individuals' lives are enormous. For example, people with mental illness reported having difficulty carrying out daily activities or requiring much energy and focus to meet demands at work [7], whereas those with depression experienced decreased enjoyment of activities and social interactions due to fluctuations in mood states [8]. Anxiety has also been found to reduce productivity and performance due to individuals' attention being excessively directed towards other people's perceptions [9]. In addition, research further demonstrated that emotional dysregulation introduces susceptibility to physical illnesses such as cardiovascular disease, viral infection, and immunodeficiency [10].

Despite the prevalence, several shortcomings exist in the current diagnosis and treatment of mental health disorders. These include comorbidities with other conditions that introduce complexity to diagnosis [11], the subsequent failure of clinicians to make accurate diagnoses due to obscurities of overlapping symptoms [11], the reliance on patients' subjective recollection of behaviors [12], and the shortage of human resources available for mental health care [13]. The limitations above contribute to underdiagnosis, preventing people in need from receiving proper treatment. In light of the need to promote more accurate detection of MH disorders, researchers began exploring the application of artificial intelligence and machine learning (ML) in this domain. Such efforts are motivated by the ability of ML to analyze large amounts of data [14], distinguish data features [15], learn meaningful relationships between data [16], and apply the identified associations to make predictions about new data [17]. Coupling ML methods with qualitative analysis, visualization, and other interpretation tools further enhances the understanding of ML outputs [17], which can support clinical decisions and improve the comprehension of causes of specific MH disorders.

Existing research has seen numerous attempts to incorporate ML in healthcare, where effective ML methods can offer automation to harness large amounts of real-time data to improve the quality of patient care [18]. Nevertheless, the dynamic nature of an individual's health, influenced by factors such as genetics, medical history, and lifestyle, remains a complex and demanding challenge to resolve [18]. Similarly, diagnoses of MH disorders are intricate due to the multifaceted nature of MH, involving emotional (e.g., sadness, helplessness), behavioral (e.g., isolation, self-talk), and physical (e.g., body aches, sleeplessness) aspects [19]. In addition, various perceived causes could contribute to MH issues, encompassing psychological (e.g., low self-esteem, overthinking), socioeconomic (e.g., racial and ethnic discrimination, poverty), and social (e.g., family conflicts, interpersonal relationships) factors [19]. As such, we hypothesize the need for multimodal data, i.e., data with multiple modalities each referring to a form of data or a signal from a data source, to achieve complementary effects for improved detection. For example, an existing work [20] has seen multimodal social media data, consisting of text, images, post metadata (e.g., time posted, likes, comments), and user metadata (e.g., profile description and image, followers), to offer additive effect when information from all modalities are incorporated. Additionally, the reliance of ML systems on extensive data and the heterogeneity of data from various sources necessitates the exploration of scalable and sophisticated ML methodologies to manage and standardize such big data, with considerations of privacy and security to ensure the confidentiality of patients' information [21].

The pipeline of ML methodologies on multimodal data includes feature extraction for each modality, transformation and fusion of modality-specific features of various structures and dimensions and ML algorithms to learn from fused representations. Our preliminary investigation of recent surveys of ML applications to multimodal MH detection revealed several data sources, such as social media [22–25], smartphones [26,27], and wearable devices [27,28]. Nevertheless, we observed a limited evaluation of the current state of knowledge in each methodological phase mentioned above, in which the understanding is crucial to inform advancements in ML approaches. In summary, the gaps we identified are the need for (1) more effective ML approaches to reduce the risk of underdiagnosis and (2) ML methodologies for handling heterogeneous and extensive multimodal data to support the detection of multifaceted MH disorders.

In this systematic literature review (SLR), we address these limitations by analyzing individual methodological phases in greater detail. We further narrow our scope to studies adopting passive sensing, which gathers users' data non-intrusively via ubiquitous sensors or devices and requires minimal user inputs. This decision is supported by our hypothesis that people's natural behaviors are best captured when their daily routines are subject to the least possible obstructions [29,30]. Less intrusive approaches have also been shown to have better acceptance among the general population, with the need to carry/wear dedicated equipment being reported as off-putting and causing levels of discomfort [12,31]. From a recent survey [17], we learned that two key motivations for ML applications to mental health are the accessibility to behavioral data enabled by continuous and non-invasive approaches and the efficiency and cost-effectiveness of timely and automated data processing. Drawing inspiration from the survey above, we establish several criteria that we anticipate in data collection approaches that are practical to promote subsequent effective detection of MH disorders: (1) reliability (i.e., ensuring that the data closely represents actual behaviors), (2) verifiable ground truth, (3) cost-effectiveness, and (4) acceptability among the general population. Consequently, we conduct a detailed analysis of each data source based on these criteria. This SLR aims to (1) assess the current trend of multimodal ML approaches for detecting various MH disorders and (2) identify an optimal strategy leveraging passively sensed multimodal data and ML algorithms. Specifically, the research questions (RQs) we aim to address throughout our study are:

- RQ1—Which sources of passive sensing data are most effective for supporting the detection of MH disorders?
- RQ2—Which data fusion approaches are most effective for combining data features of varying modalities to prepare for training ML models to detect MH disorders?
- RQ3—What ML approaches have previous researchers used to successfully detect MH disorders from multimodal data?

The SLR is structured as follows: Section 2 outlines the research methods adopted in this review, followed by Section 3, which presents results that analyze the individual phases of existing methodologies, including data sources, feature extraction, modality fusion techniques, and the ML algorithms adopted. Based on the analysis, Section 4 then synthesizes the findings to address each RQ mentioned above and draws insights into recommendations and considerations for future researchers wishing to innovate in this space. Lastly, Section 5 concludes the study.

2. Materials and Methods

This section presents the review protocol for our SLR based on the PRISMA 2020 Statement [32], a guideline for healthcare-related studies [33] established based on the PRISMA 2009 Statement [34], and the Guidelines for SLRs in Software Engineering [35] published in 2007.

2.1. Search Strategy

We performed an exhaustive search on four online databases: Scopus, PubMed, ACM Digital Library, and IEEE Xplore. We chose these databases due to the abundance of published papers on the topic of concern and to represent the multidisciplinarity of the topic by having a diversity of papers across the fields of clinical science and computing science. As previously explained, we concentrate on studies utilizing data of at least two different modalities collected using ubiquitous devices and applying ML techniques for detecting MH disorders. Inspired by Zhang et al.'s [36] search strategy, we systematically constructed our search query based on aspects shown in Table 1.

We queried the databases by combining keywords within the same category with an OR operator and those across categories with an AND operator. We also considered different terminology variants by using wildcards (*), for instance, "well*" in our query string, because the term "wellbeing" may be spelled as "well-being" in certain studies. An example of our query string on Scopus is as follows:

"ALL (mental AND (health OR disorder OR illness OR well*)) AND TITLE-ABS-KEY ("artificial intelligence" OR "machine learning" OR model) AND TITLE-ABS-KEY (detect* OR predict* OR classif* OR monitor* OR recogn* OR identif*) AND TITLE-ABS-KEY ("social media" OR text* OR audio* OR speech* OR voice OR visual OR imag* OR video* OR smartphone* OR mobile OR wearable* OR sens*) AND PUBYEAR > 2014".

Table 1. Search categories and keywords.

Category	Keywords
Mental disorder	Mental health, mental disorder, mental illness, mental wellness, mental wellbeing
Method	Artificial intelligence, machine learning, model
Outcome	Detect, predict, classify, monitor, recognize, identify
Data source/modality	Social media, text, speech, voice, audio, visual, image, video, smartphone, mobile, wearable, sensor

We decided on the cutoff publication year of 2015 due to the consideration of the developmental trajectory of the research domain. Our preliminary observation revealed 2015 as a potential juncture where relevant studies began gaining momentum, coinciding with the introduction of AVEC 2013 [37] and AVEC 2014 [38] challenges focusing on facial expressions and vocal cues relating to specific MH conditions such as depression. We intended to ensure that our review encompasses more recent advancements for a comprehensive understanding of the field's current state. In addition, the rapid development of technologies may render techniques from older publications obsolete or less relevant. Likewise, the relevance of findings related to smartphones and wearable devices may have evolved due to changes in their adoption among the general population over time.

2.2. Inclusion and Exclusion Criteria

To ensure the selection of studies that align with our research focus, we considered a study to be relevant if it fulfilled all of the following inclusion criteria:

- The study collects data passively via ubiquitous or wearable devices, considering the cost-effectiveness and general accessibility.
- The data is human generated, i.e., derived from individuals' actions in an environment or interactions with specific platforms or devices.
- The data source involves at least two different modalities.
- The study adopts ML algorithms intending to detect one or more MH disorders.
- The study is written in English.
- The study was published from the year 2015 onwards (further details in the following section).

We excluded a study if any of the following exclusion criteria were satisfied:

- The study investigates data sources of a single modality or exclusively focuses on a specific modality, e.g., text-based approaches.
- The study specifically targets the pediatric population, i.e., toddlers and children below ten years old, as defined within the suggested adolescent age range of 10–24 years [39].
- The study targets a particular symptom of specific MH disorders, e.g., low mood, which is a common sign of depression.
- Data collection requires dedicated equipment or authorized resources:
    - Brain neuroimaging data, e.g., functional magnetic resonance imaging (fMRI), structural MRI (sMRI), electroencephalogram (EEG), electromyography (EMG), and photoplethysmography (PPG) signals
    - Clinical data, e.g., electronic health records (EHRs) and clinical notes
    - Genomic data
    - Body motions collected using specialized motion capture platforms or motor sensors
    - Makes use of Augmented Augmented Reality (AR) or Virtual Reality (VR) technology
- The study does not employ ML algorithms for detection/prediction, e.g., focusing on correlation/association analysis, treatment/intervention strategies, or proposing study protocols.
- The study is a survey, book, conference proceeding, workshop, or magazine

- The study is unpublished or non-peer-reviewed.

Since our work explicitly emphasizes multimodality to observe cross-modality fusion and interactions, we excluded studies emphasizing a single modality. For example, we do not consider those solely analyzing textual content from social media sources (e.g., Twitter) without incorporating broader online social behaviors, such as posting time distribution and interactions with other users through retweets and comments. Additionally, we omitted studies involving children, as it is well-established that factors, manifestations, and responses to MH conditions can differ significantly between children and adults [40]. Children may also often rely on parents and family environment for care and treatment [41,42].

While changes in MH states such as affect, emotion, and stress may serve as potential indicators of MH disorders such as depression and anxiety [43], it is noteworthy that these factors, when considered in isolation, do not necessarily equate to a complete MH diagnosis [44,45]. Therefore, we refined our focus by excluding studies that solely investigated these states. Due to practicality concerns, we also enforced the utilization of ubiquitous devices in data collection to ensure these tools are easily accessible and cost-effective.

2.3. Selection Process

Figure 1 shows the literature search process as a flow diagram adapted from an example in the PRISMA guideline (https://www.bmj.com/content/bmj/339/bmj.b2700/F2.large.jpg (accessed on 19 September 2023)). After querying the selected databases, we re-evaluated the title, abstract, and keywords of individual studies to refine the results and remove duplicates. Subsequently, we manually applied the eligibility criteria to determine relevant studies for data extraction.

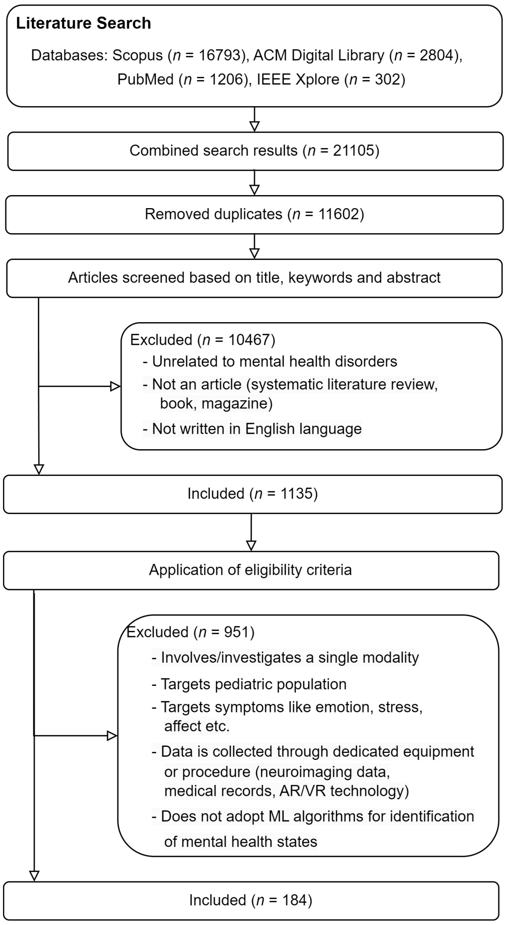
Literature Search
Databases: Scopus (n = 16793), ACM Digital Library (n = 2804), PubMed (n = 1206), IEEE Xplore (n = 302)
Combined search results (n = 21105)
Removed duplicates (n = 11602)
Articles screened based on title, keywords and abstract
Excluded (n = 10467)
- Unrelated to mental health disorders
- Not an article (systematic literature review, book, magazine)
- Not written in English language
Included (n = 1135)
Application of eligibility criteria
Excluded (n = 951)
- Involves/investigates a single modality
- Targets pediatric population
- Targets symptoms like emotion, stress, affect etc.
- Data is collected through dedicated equipment or procedure (neuroimaging data, medical records, AR/VR technology)
- Does not adopt ML algorithms for identification of mental health states
Included (n = 184)

Figure 1. Flow diagram of study selection.

2.4. Data Extraction

Table 2 shows the information we extracted from individual studies and the corresponding mapping to the relevant research questions (RQs) where applicable.

Table 2. Data to extract to answer respective research questions.

ID	Item	RQ
I1	Reference (authors and year)	N/A
I2	Title	N/A
I3	Mental health disorder investigated	N/A
I4	Data collection process	RQ1
I5	Ground truth/data labeling	RQ1
I6	Feature extraction process	RQ2
I7	Feature transformation process if any	RQ2
I8	Feature fusion process	RQ2
I9	Machine learning model	RQ3
I10	Results achieved	N/A
I11	Analysis findings if any	N/A

2.5. Quality Assessment

We adapted a suggested checklist [35] to develop quality assessment criteria, shown in full in Table 3, that assigns a score to each study.

Table 3. Quality assessment criteria and scoring.

ID	Criteria	Scoring
QC1	Was there an adequate description of the context in which the research was carried out?	The design, setup, and experimental procedure are adequately (1), partially (0.5), or poorly described (0)
QC2	Were the participants representative of the population to which the results will generalize?	The participants fully (1), partially (0.5), or do not (0) represent the stated target population
QC3	Was there a control group for comparison?	Control group has (1) or has not (0) been included
QC4	Were the measures used in the research relevant for answering the research questions?	Adopted methodology and evaluation methods are fully (1), partially (0.5), or not (0) aligned with research objectives
QC5	Were the data collection methods adequately described?	Data collection methods are adequately (1), partially (0.5), or poorly (0) described
QC6	Were the data types (continuous, ordinal, categorical) and/or structures (dimensions) explained?	All (1), some (0.5), or none (0) of the data types and structures of various modalities are explained
QC7	Were the feature extraction methods adequately described?	Feature extraction methods are adequately (1), partially (0.5), or poorly (0) described
QC8	Were the machine learning approaches adequately described?	Machine learning models and architectures are adequately (1), partially (0.5), or poorly (0) described
QC9	On a scale of 1–5, how reliable/effective was the machine learning approach?	Effectiveness, reliability and consistency of machine learning approach is well (5), partially (3), or poorly (0) justified through evaluation, analysis and baseline comparison
QC10	Was there a clear statement of findings?	Experimental findings are well (1), partially (0.5), or poorly (0) described
QC11	Were limitations to the results discussed?	Result limitations are well (1), partially (0.5), or poorly (0) identified
QC12	Was the study of value for research or practice?	Research methodology or outcomes well (1), partially (0.5), or poorly (0) contribute valuable findings or application

Most scoring, except for QC3, QC9, and QC13, adopt a three-item scale, satisfies = 1, does not satisfy = 0, and partially satisfies = 0.5, to evaluate whether a study complies with the corresponding criteria. The final quality score would be the summation of the score corresponding to the conformity of the checklist items. Acknowledging that healthy controls are not always necessary in relevant studies, we have specifically included checklist item QC3 due to our interest in the effectiveness of data sources and methodological approaches in distinguishing between individuals with and without MH disorders. Understanding general patterns in healthy controls also serves as a baseline for benchmarking to justify the significance of future findings related to those with MH conditions.

3. Results

This section summarizes and analyzes the results we extracted from 184 relevant studies published from January 2015 to August 2023 based on Table 2. Table 4 displays the combinations of MH conditions investigated and the categories of data sources involved in all selected studies. Figure 2 shows the methodological pipeline involved in our data extraction. The following subsections describe and explain each extracted finding in detail.

Table 4. A compilation of relevant studies for data extraction.

Mental Health Conditions	Data Source
Depression	AV [43,46–108] SM [20,25,98,109–148] SS [99,100,104,105,149–177] WS [149–151,155–158,164,169,171–173,178–182]
Suicidal intent	AV [100,183,184] SM [185–189] SS [100,147,190] WS [181,182,191]
Bipolar disorder	AV [101–103,192–200] SM [201] SS [12,172,200,202] WS [172]
Schizophrenia	AV [203] SM [201,204] SS [172,205–209] WS [172,210]
Anxiety	AV [104,105,108,211] SM [148] SS [104,105,173–176,212] WS [173,179,213]
Autism spectrum disorder	AV [214] WS [215]
PTSD	AV [107,216] WS [216]
Eating disorder	SM [217–219]
Mental illness	SM [220]
Mental wellbeing	AV [221] SS [222,223]

AV: Audio and video recordings, SM: Social media, SS: Smartphone sensors, WS: Wearable sensors.

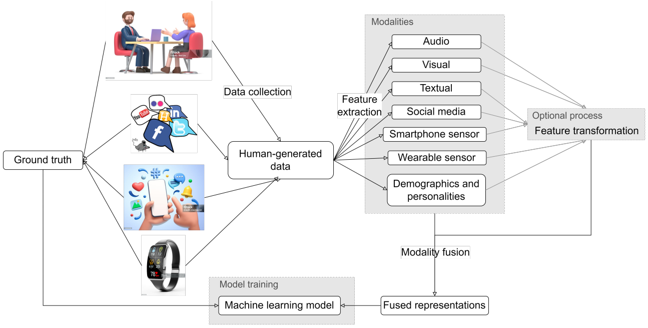

Figure 2. Pipeline of methodological phases involved in data extraction [224–227].

3.1. Data Source

The primary categories of data sources are (1) audio and video recordings (n = 82), (2) social media (n = 55), (3) smartphones (n = 54), and (4) wearable devices (n = 28).

3.1.1. Audio and Video Recordings

Audio and video recordings of individuals were captured using video cameras, webcams, or microphones while they responded to interview questions or completed predetermined tasks in person or online. For example, Gratch et al. [228] conducted semi-structured interviews with individual participants with both neutral questions and those related to depression or PTSD events, which the authors recorded using a camera and close-talking microphone. In contrast, NEMSI (NEurological and Mental health Screening Instrument) [229] was proposed as a cloud-based system that automates data capture and the subsequent audio-visual processing for feature extraction and visualization. Before commencing the interviews, researchers [228,230] ensured that participants signed consent forms to collect highly identifiable recording data and share their data for research purposes. The researchers also offered transparency regarding the purpose of their study and data collection before participants provided their consent.

3.1.2. Social Media

Meanwhile, social media platforms like Twitter, Reddit, Sina Microblog, Instagram, Facebook, YouTube, Flickr, and Blued offer a safe space for information sharing, communication, and expressing emotions. Various forms of user-generated content publicly available on these platforms are texts, images, social interactions (likes, comments, mentions, and shares), and user profile information (followers, followings, bio descriptions, profile images). Researchers could crawl content from these platforms using the provided application programming interface (API) by strategically querying content posted within a predetermined duration for observation, locating the presence of relevant phrases or keywords within textual content, or sourcing directly from discussion space revolving around specific MH conditions where applicable. For instance, Shen et al. [20] identified candidate social media users based on tweets containing the character string "depress" and utilized such tweets as anchor points to sample remaining tweets posted by the corresponding users within a month relative to anchor tweets. Meanwhile, Mishra et al. [185] scraped the top 100 posts from the "r/suicidalthoughts", "r/suicidewatch", and "takethislife.com" forums with an abundance of posts related to suicidal ideation.

Nevertheless, we observed limited ethical considerations and explicit mentions in existing studies regarding obtaining participants' consent for utilizing their data for research purposes. For example, Yates et al. [231] discussed the privacy risks with posts crawled from Reddit as minimal since this data is publicly available on the platform. The researchers also described their privacy measures for ensuring that annotators and other researchers were only allowed access to anonymized posts after agreeing to adhere to the ethical guidelines for not attempting to contact or deanonymize data samples.

3.1.3. Smartphones

Smartphone sensors, such as accelerometers, GPS, light sensors, and microphones, could collect and infer information about smartphone usage, physical activity, location, and an individual's environment. Researchers have adopted existing mobile applications that collect sensing data, such as Purple Robot [232] (Android only), SensusMobile [99] (Android only), and LifeRhythm [233] (Android and iOS), and those with additional features, including Behavidence https://www.behavidence.com/ (accessed on 10 December 2023) (Android application that displays similarity scores of inferred behaviors to specific MH disorders), Insights [234] (Android application with customizable questionnaires), MoodMirror [235] (Chinese Android application that connects with a wristband via Bluetooth), or BiAffect https://www.biaffect.com/ (accessed on 10 December 2023) (iOS only), that collect keyboard typing data specifically. In contrast, some researchers developed mobile applications for their use cases using frameworks like AWARE [236] (collects sensor data from Android and iOS devices and supports integration with data analysis pipeline). These mobile applications act as a central management system, either storing data locally in individuals' devices or transmitting them to a central server for processing and analysis.

Prior to data collection, researchers obtained participants' consent and provided details about the data to be collected. Some researchers additionally conducted onboarding sessions for installing mobile applications, offered tutorials to operate them, and provided technical support throughout the data gathering duration [237]. Privacy measures were also implemented to minimize identifiability and the risks of data leakage during transmission, such as anonymizing participants, hashing phone calls and text messaging logs [237], and employing secure transmission protocols like HTTPS and SSL.

3.1.4. Wearable Devices

Wearable devices have further enabled the collection of physical activity, movement, sleep, and physiological signals like heart rate (HR), electrodermal activity (EDA), skin temperature (ST), and galvanic skin response (GSR). Some examples of wearables are Empatica E4 wristbands [149,172,178], Microsoft Band 2 [150], Fitbit Charge or Flex trackers [151,155,164,180–182,191], and the Galaxy S3 smartwatch [169]. Data gathered through these devices were transmitted directly to an internet-connected server [215] or transferred via Bluetooth [210] to dedicated mobile applications that handle the transmission as described above. As such, existing studies executed similar procedures for obtaining participants' consent before data collection and privacy measures to ensure secure data transmission.

Table 5 describes publicly available datasets discovered or released by studies included in this work for multimodal detection.

Table 5. Public datasets and respective data source categories.

Dataset	Description	Mental Health Disorders	Source Category
Distress Analysis Interview Corpus—Wizard of Oz (DAIC-WOZ) [228]	Video recordings and text transcriptions of interviews conducted by a virtual interviewer on individual participants (used in Audio-Visual Emotion Challenge (AVEC) 2014 [38], 2016 [47], 2017 [238], and 2019 [239])	Post-traumatic stress disorder (PTSD), depression, anxiety	AV
Turkish Audio-Visual Bipolar Disorder Corpus [230]	Video recordings of patients during follow-ups in a hospital	Bipolar disorder	AV
Engagement Arousal Self-Efficacy (EASE) [240]	Video recordings of individuals undergoing self-regulated tasks by interacting with a website	PTSD	AV
Well-being [241]	Video recordings of conversational interviews conducted by a computer science researcher	Depression, anxiety	AV
Emotional Audio-Textual Depression Corpus (EATD-Corpus) [73]	Audio responses and text transcripts extracted from student interviews conducted by a virtual interviewer through an application	Depression	AV
Reddit Self-Reported Depression Diagnosis Corpus (RSDD) [231]	Reddit posts of self-claimed and control users	Depression	SM
Self-Reported Mental Health Diagnosis Corpus (SMHD) [242]	Twitter posts of users with one or multiple mental health conditions and control users	ADHD, anxiety, autism, bipolar disorder, borderline personality disorder, depression, eating disorder, OCD, PTSD, schizophrenia, seasonal affective disorder	SM

Table 5. *Cont.*

Dataset	Description	Mental Health Disorders	Source Category
Multi-modal Getty Image depression and emotion (MGID) dataset [106]	Textual and visual documents from Getty Image with equal amount of depressive and non-depressive samples	Depression	SM
Sina-Weibo suicidal dataset [243]	Sina microblog posts of suicidal and control users	Suicidal ideation	SM
Weibo User Depression Detection dataset (WU3D) [112]	Sina microblog posts of depressed candidates and control users, and user information such as nickname, gender and profile description	Depression	SM
Chinese Microblog depression dataset [244]	Sina microblog posts following the last posts of individuals who have committed suicide	Depression	SM
eRisk 2016 dataset [245]	Textual posts and comments of depressed and control users from Twitter, MTV’s A Thin Line (ATL) and Reddit	Depression	SM
eRisk 2018 dataset [246]	Textual posts and comments from Twitter, MTV’s A Thin Line (ATL) and Reddit	Depression, anorexia	SM
StudentLife [237]	Smartphone sensor data of students from a college	Mental wellbeing, stress, depression	SS
CrossCheck [205]	Smartphone sensor data of schizophrenia patients	Schizophrenia	SS
Student Suicidal Ideation and Depression Detection (StudentSADD [100]	Voice recordings and textual responses obtained using smartphone microphones and keyboards	Suicidal ideation, depression	AV, SS
BiAffect dataset [247]	Keyboard typing dynamics captured by a mobile application	Depression	SS
Tesserae dataset [248]	Smartphone and smartwatch sensor data, Bluetooth beacon signals, and Instagram and Twitter data of information workers	Mood, anxiety, stress	SS, WS, SM
CLPsych 2015 Shared Task dataset [249]	Twitter posts of users who publicly stated a diagnosis of depression or PTSD with corresponding control users of the same estimated gender with the closest estimated age	Depression, PTSD	SM
multiRedditDep dataset [128]	Reddit images posted by users who posted at least once in the */r/depression* forum	Depression	SM
Fitbit Bring-Your-Own-Device (BYOD) project by “All of Us” research program [250]	Fitbit data (e.g., steps, calories, and active duration), clinical assessments, demographics	Depression, anxiety	WS
PsycheNet dataset [138]	Social contagion-based dataset containing timelines of Twitter users and those with whom they maintain bidirectional friendships	Depression	SM

Table 5. *Cont.*

Dataset	Description	Mental Health Disorders	Source Category
PsycheNet-G dataset [139]	Extends PsycheNet dataset [138] by incorporating users' social interactions, including bidirectional replies, mentions, and quote-tweets	Depression	SM
Spanish Twitter Anorexia Nervosa (AN)-related dataset [251]	Tweets posted by users whom clinical experts identified to fall into categories of AN (at early and advanced stages of AN but do not undergo treatment), treatment, recovered, focused control (control users that used AN-related vocabulary), and random control	AN	SM
Audio-visual depressive language corpus (AViD-Corpus) [37]	Video clips of individuals performing PowerPoint-guided tasks, such as sustained vowel, loud vowel, and smiling vowel phonations, and speaking out loud while solving a task (used in AVEC 2013 [37])	Depression	AV
Existing call log dataset [222]	Call and text messaging logs and GPS data collected via mobile application and in-person demographic and mental wellbeing surveys	Mental wellbeing	SS
Speech dataset [252]	Audio recordings of individuals performing two speech tasks via an external web application and demographics obtained from recruitment platform, Prolific [253]	Anxiety	AV
Early Mental Health Uncovering (EMU) dataset [104]	Data gathered via a mobile application that collects sensor data (i.e., text and call logs, calendar logs, and GPS), Twitter posts, and audio samples from scripted and unscripted prompts and administers PHQ-9 and GAD-7 questionnaires and demographic (i.e., gender, age, and student status) questions	Depression, anxiety	SS, AV, SM
Depression Stereotype Threat Call and Text log subset (DepreST-CAT) [105]	Data gathered via modifying the EMU application [104] to collect additional demographic (i.e., gender, age, student status, history of depression treatment, and racial/ethnic identity) and COVID-19 related questions	Depression, anxiety	SS, AV, SM
D-vlog dataset [92]	YouTube videos with equal amounts of depressed and non-depressed vlogs	Depression	AV

AV: Audio and video recordings, SM: Social media, SS: Smartphone sensors, WS: Wearable sensors.

3.2. Data Ground Truth

Data used for supervised learning must have a ground truth (i.e., if the person to whom the data belong suffers from a specific MH disorder) so that ML models learn to distinguish data points of different ground-truth labels. The means of ground truth acquisition are (1) clinical assessment by trained psychiatrists or healthcare professionals and (2) self-reports by people themselves.

3.2.1. Clinical Assessments

During clinical diagnoses, trained psychiatrists use clinically validated assessment scales with known symptoms of specific MH disorders to prompt patients to share their experiences. Establishing ground-truth knowledge varies based on experimental design in existing studies, where trained healthcare professionals could conduct clinical assessments before the data collection procedure and during other intermediate phases deemed necessary. For example, Grünerbl et al.'s [12] study involved psychologists conducting examinations every three weeks over the phone, using standard scale tests such as the Hamilton Rating Scale for Depression (HAMD) or Young Mania Rating Scale (YMRS), whereas participants were scheduled for monthly face-to-face clinical assessments with clinicians using the 7-item Brief Psychiatric Rating Scale (BPRS) in Wang et al.'s [206] study. On the other hand, participants could be recruited from the MH service within a hospital setting, where existing diagnoses of specific MH conditions are known, and clinical assessments could be reconducted during follow-ups and after discharge using the YMRS [230].

If access to healthcare professionals is unavailable, these scales can be administered through mobile applications or other devices to be answered and self-reported by subjects. Examples of scales used in both clinical and self-reported assessments are the Hamilton Depression Rating Scale (HDRS) [254], Patient Health Questionnaire-9 (PHQ-9) [255], Beck Depression Inventory (BDI) [256], and Center for Epidemiological Studies Depression Scale (CES-D) [257]. Researchers can compile and analyze the responses to derive ground truth based on established guidelines. For example, the summation score of the PHQ-9 scale corresponds to depression severity levels, where 5, 10, 15, and 20 represent mild, moderate, moderately severe, and severe depression, respectively [258].

3.2.2. Self-Reports

In most cases where social media data has been scraped from public-facing platforms via application programming interfaces, users are not reachable due to security and privacy protection. As such, their MH states are not immediately acquirable since they do not usually disclose invasive information like medical history. Researchers have relied on textual or visual cues in users' public posts to locate the existence of MH disorders for the purposes of ground truth. They detected self-reports where users explicitly disclosed being diagnosed with a specific MH disorder in their public posts by looking for sentence structure such as "I am diagnosed with ..." [20].

This ground-truth acquisition method heavily relies on individuals' willingness and openness to share content publicly on social media platforms. Therefore, to enhance the accuracy of ground truth labels, studies incorporated clinical opinions when annotating and labeling social media data. These opinions were sourced from trained psychiatrists or psychologists [112,131,145,186,219], as well as staff and students within the university settings with backgrounds in psychology [142,185]. For example, Abuhassan et al. [218] incorporated opinions from domain experts with specific expertise in eating disorders (EDs), psychology, mental health, and social media. The authors obtained a comprehensive and well-rounded annotation strategy to guide the categorization of social media users into individuals with an explicit diagnosis of EDs, healthcare professionals, communicators (i.e., those who communicate, exchange, and distribute information to the public), and non-ED individuals. The approaches above attempted to address the possibility of researchers overlooking implicit indicators of specific MH disorders or lacking sufficient clinical knowledge to make accurate inferences based on several posts created by each individual [135,189]. However, these efforts may not suffice, given that public content posted by individuals might be adapted with considerations of self-presentation factors.

3.3. Modality and Features

A range of modality-specific features within the datasets analyzed by researchers were found to support the identification of MH-related features in study participants. Table 6 provides a summary of these features and their findings relevant to MH diagnosis. See Appendix A for a more extensive view of the features and the corresponding extraction tools.

Table 6. Categories of modality features.

Modality	Category	Description	Examples
Audio	Voice	Characteristics of audio signals	Mel-frequency cepstral coefficients (MFCCs), pitch, energy, harmonic-to-noise ratio (HNR), zero-crossing rate (ZCR)
	Speech	Speech characteristics	Utterance, pause, articulation
	Representations	Extracted from model architectures applied onto audio samples or representations	Features extracted from specific layers of pre-trained deep SoundNet [259] network applied onto audio samples
	Derived	Derived from other features via computation methods or models	High-level features extracted from long short-term memory (LSTM) [260] model applied onto SoundNet representations to capture temporal information
Visual	Subject/object	Presence or features of a person or object	Face appearance, facial landmarks, upper body points
	Representations	Extracted from model architectures applied onto image frames or representations	Features extracted from specific layers of VGG-16 network [261] (pre-trained on ImageNet [262]) applied onto visual frames
	Emotion-related	Capture emotions associated with facial expressions or image sentiment	Facial action units (FAUs) corresponding to Ekman's model of six emotions [263], i.e., anger, disgust, fear, joy, sadness, and surprise, or eight basic emotions [264] that additionally include trust, negative and positive
	Textual	Textual content or labels	Quotes in images identified via optical character recognition (OCR)
	Color-related	Color information	Hue, saturation, color
	Image metadata	Image characteristics and format	Width, height, presence of exchangeable image file format (exif) file
	Derived	Derived from other features via computation methods or models	Fisher vector (FV) encoding [265] of facial landmarks
Textual	Linguistic	Language in terms of choice of words and sentence structure	Pronouns, verbs, suicidal keywords
	Sentiment-related	Emotion and sentiment components extracted via sentiment analysis (SA) tools	Valence, arousal and dominance (VAD) ratings
	Semantic-related	Meaning of texts	Topics and categories describing text content
	Representations	Vector representations generated using language models	Features extracted from pre-trained Bidirectional Encoder Representations from Transformers (BERT) [266] applied onto texts
	Derived	Derived from other features via computation methods or models	Features extracted from LSTM with attention mechanism applied onto textual representations to emphasize significant words
Social media	Post metadata	Information associated with a social media post	Posting time, likes received
	User metadata	Information associated with a social media user account	Profile description and image, followers, followings
	Representations	Representations of social network and interactions with other users	Graph network representing each user using a node and connecting two users mutually following each other
	Derived	Derived from other features via aggregation or encoding	Number of posts made on the weekends

Table 6. *Cont.*

Modality	Category	Description	Examples
Smartphone sensor	Calls and messages	Relating to phone calls and text messaging	Frequency and duration of incoming/outgoing phone calls
	Physical mobility	Inferences from accelerometer, gyroscope, and GPS data	Walking duration, distance traveled
	Phone interactions	Accessing phone, applications, and keyboards	Duration of phone unlocks, frequency of using specific applications, keystroke transitions
	Ambient environment	Surrounding illumination and noise	Brightness, human conversations
	Connectivity	Connections with external devices and environment	Association events with WiFi access points, occurrences of nearby Bluetooth devices
	Representations	High-level representations of time series sensor data	Features extracted from transformer to capture temporal patterns
	Derived	Derived from low-level features via computation or aggregation	Average weekly visited location clusters, sleep duration estimated from phone being locked and being stationary in a dark environment at night
Wearable sensor	Physical mobility	Inferences related to physical motion and sleep	Number of steps, sleep duration and onset time
	Physiological	Physiological signals	Heart rate, skin temperature
	Representations	High-level representations of time series sensor data	Features extracted from LSTM applied onto heart rate signals
Demographics and Personalities	Demographic	Personal demographic information	Age, gender
	Personality	An individual's personality	Big 5 personality scores

3.3.1. Audio

Several popular approaches to extract audio features include adopting OpenSmile [267] to extract low-level descriptors (LLDs) and employing pre-trained deep learning (DL) models to extract high-level deep representations from either audio samples directly or transformed spectrogram images [66]. Researchers have identified several audio features to be significant indicators of MH conditions. For instance, Yang et al. [193] discovered histogram-based audio LLDs to be more effective than visual features in identifying bipolar disorder, and such indicators are more prominent in male samples. Meanwhile, from specific features such as energy contours, kurtosis, skewness, voiced tilt, energy entropy, and MFCCs, Belouali et al. [184] demonstrated that individuals with suicidal intent spoke using a less animated voice with flatter energy distribution and fewer bursts. Their speech had less vocal energy and less abrupt changes and were more monotonous. Other audio features found to be significant indicators of depression and PTSD include audio intensity, pitch, and spectral decrease [54,107]. Since it is beyond the scope of the current work to dive deep into audio samples and features, we direct interested researchers to an existing work [268] for greater details on audio processing and features that could be extracted at varying domains (e.g., time, frequency, and cepstrum) [54].

3.3.2. Visual

Visual features were extracted by first locating individuals or objects in a video frame or a static image, identifying the corresponding feature points (e.g., facial landmarks, FAUs, upper body points), and then generating features using image processing tools including OpenFace [269], OpenCV [270], and OpenPose [271]. Pre-trained models could also be applied directly to visual samples to extract feature representations. For image frames extracted from video recordings, researchers could further capture dynamic aspects and transitions across a video, such as computing the speed and range of displacements of specific feature points between successive video frames and the variation across the entire video.

Facial action units (FAUs) were introduced to describe facial movements [272], where each AU corresponds to contractions of specific facial muscles (e.g., AU5 represents raised upper eyelids, AU6 represents raised cheeks, and AU15 represents pulled-down lip corners [273] as shown in Figure 3). FAUs have shown significant promise in encoding facial expressions, each constituted by a combination of AUs [273] as shown in Figure 4. In the current context, Thati et al. [99] demonstrated that a few AUs correlate significantly with depressive symptoms, specifically, AU12, AU10, and AU25, corresponding to pulled lip corners, raised upper lips, and parted lips, respectively. Referring to both Figures 3 and 4, this finding could be associated with a smiling expression comprising AU12 and AU25 and low mood as demonstrated by AU10. It could potentially indicate the "smiling depression" scenario mentioned by Ghosh et al. [132], where individuals with depression may choose to post more happy images compared to healthy controls who expressed diverse emotions.

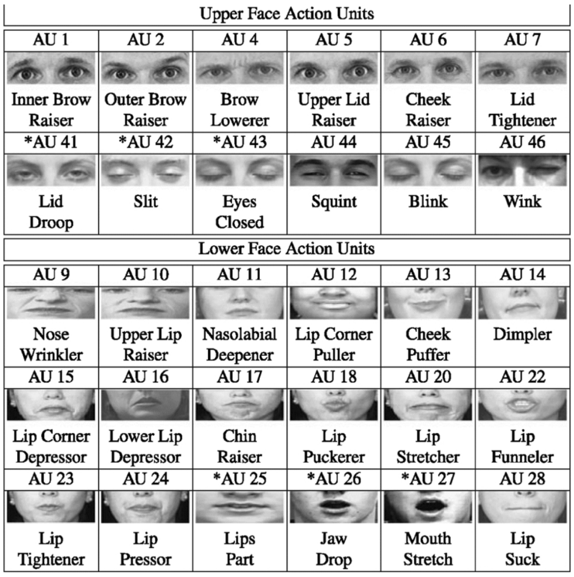

Figure 3. Examples of facial movements coded using facial action units [274].

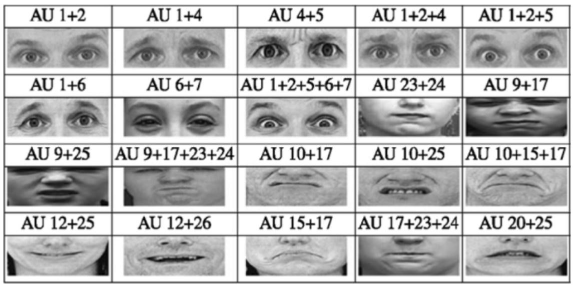

Figure 4. Examples of facial expressions resulting from combinations of facial action units [274].

In addition, facial appearance and emotions in shared images were significantly indicative of depression and PTSD [107,116]. While a few studies [25,132,220] showed that individuals with depression have lower tendencies to disclose facial identity, Gui et al. [111] found that they are more likely to post images with faces but of a lower average face count per image. From the revelation of more images of animals and health objects from Twitter and Reddit content, Uban et al. [128] hypothesized the possibility of individuals' online help-seeking through viewing animal-related content that might improve psychological and emotional conditions and looking up causes of health events, diseases, and treatment options.

Meanwhile, other research studies [25,111,220] revealed that individuals with mental illness and depression post less colorful images of darker and grayer colors on social media compared to healthy controls, who prefer brighter and more vivid colors such as blue and green. These patterns potentially align with existing knowledge [275] regarding the influence of individuals' mood on color preferences, where principal hues (e.g., red, yellow) and intermediate hues (e.g., yellow-red, blue-green) evoked higher positive emotions than achromatic colors like black and white. Specifically, Yazdavar et al. [25] demonstrated a strong positive correlation between self-reported depressive symptoms and individuals' tendency to perceive surroundings as grey or lacking colors. In contrast, Xu et al. [220] further computed pleasure, arousal, and dominance scores from brightness and saturation values. The authors then discovered that individuals with mental illness preferred less saturated images (i.e., containing more grey [276]), which implied higher dominance and arousal than healthy controls.

3.3.3. Textual

In addition to textual content written by individuals, researchers also obtained textual transcripts from audio samples using speech-to-text tools on Google Cloud Platform, AWS Transcribe [277], or Transcriber-AG https://transag.sourceforge.net/ (accessed on 10 December 2023). Tools like Linguistic Inquiry and Word Count (LIWC) [278], Suite of Automatic Linguistic Analysis Tools (SALAT) [279], and Natural Language Toolkit (NLTK) [280] were adopted on textual content to identify nouns, adjectives, pronouns, or specific words referring to social processes and psychological states, where linguistic features were generated as the occurrence of words in specific categories. Meanwhile, sentiment-related features like sentiment polarity scores were obtained from sentiment analysis tools, including Stanford NLP toolkit [281], Sentiment Analysis and Cognition Engine (SEANCE) [282], and Affective Norms for English Words ratings (ANEW) [283]. High-level textual representations could also be obtained via language models, such as BERT [266], Paragraph Vector (PV) [284], and XLNet [285], to represent each word using a vector. The overall textual representations could be obtained via concatenating directly, averaging, or applying attention mechanisms on word-level representations to emphasize more significant features.

Abundant studies consistently highlighted the prominent correlations between textual features and MH conditions. For example, the significance of linguistic features in MH identification was accentuated by compelling evidence showing that individuals with depression and suicidal intent used more first-person pronouns, possibly reflecting their suppressed nature. This linguistic pattern was observed in textual content across various social media platforms, including Weibo [109,123], Instagram [116], Twitter [25,118,121], and Reddit [186], as well as in transcribed audio recordings [184]. Meanwhile, several other studies [123,187] further found more frequent usage of the word "others" or third-person pronouns (e.g., "they", "them", "he", "she") than healthy controls, which the authors hypothesized as the tendency of depressive or suicidal individuals in acquiring physiological distance and reluctant to show feelings. In addition, researchers found individuals with depression, suicidal intent, and schizophrenia exhibiting a pronounced expression of negative emotions compared to healthy controls. This observation is substantiated by various features, including the frequency of negative words [20,118,123,204,220] and negative emoticons [130,188], as well as negative sentiment scores of overall sentences [121].

In contrast, specific keywords could be indicative, such as references to personal events like "work pressure", "divorce", and "break up" [118]; biological processes like "eat", "blood", and "pain" [116]; or family references like "daughter", "dad", and "aunt" [184]. Existing studies also revealed keywords or phrases related to specific MH conditions to be helpful. For example, in depression detection, researchers [54] identified prominent usage of words associated with depressive symptoms, such as "depressed", "hopeless", and "worthless", referenced from the Depression Vocabulary Word List https://myvocabulary.com/word-list/depression-vocabulary/ (accessed on 10 December 2023), as well as antidepressant names [123] based on the list from Wikipedia https://en.wikipedia.org/wiki/List_of_antidepressants (accessed on 10 December 2023). Nevertheless, MH-related keywords may be expressed differently for various MH disorders. For instance, individuals with schizophrenia used more words related to perception (hear, see, feel), swearing, and anger [204], while Tébar et al. [217] found individuals with an eating disorder (ED) publishing less ED-related content, involving fewer indicative terms like "laxative names" and "weight concerns", to keep their illness private. The latter study demonstrated false positives introduced by such disparities, as healthy controls were involved in discussions of MH disorders like PTSD or depression that share some ED symptoms or mentioned prevalent topics in the pro-ED community.

3.3.4. Social Media

On top of user-generated texts and images on social media platforms, researchers could infer social networks and interactions from metadata associated with users and posts, where followers and followings could indicate "friendships", whereas interactions like posting, liking, and commenting could reveal social interactions and topics of interest. While most platforms offer fundamental post information such as time posted, likes, and comments, some details are platform-specific, such as retweets (Twitter), check-in locations (Facebook), favorites (Twitter), profile images (Instagram), and users' details like age and gender (Sina Microblog). Graph architectures could then be adopted to model the information above, for instance, by having a node for each user and an edge between two nodes representing the presence or extent of particular social interactions.

Research attempts have demonstrated a significant association between time spent on social media platforms [116] and depressive symptoms. This claim is supported by compelling evidence indicating that a substantial proportion of individuals with depression (76% [130]) and suicidal intent (73% [188]) engaged in more active posting activities on various social media platforms, including Instagram [125], Twitter [20,118], Reddit [130,188], and Weibo [189], particularly at midnight. Some authors [20,118,134] intuited this behavior as potentially linked to sleeping problems or insomnia. The corresponding posts by these individuals were also found to receive less engagement and attention, such as likes, retweets, and favorites [122,137]. Nevertheless, researchers observed contradicting trends in posting behaviors. While a few studies [122,125,137,189] revealed that those with MH disorders were generally less active on Twitter and Instagram, the opposite was observed in other studies on Twitter [121] and Sina Microblog [109]. Such disparities could be attributed to different user populations or sampling periods that may influence social behaviors on these platforms. Other potentially depressive behaviors include less disclosure of personal information [123], a greater likelihood of modifying images before posting [220], and lower preferences for sharing location [122].

Several research attempts emphasized the role of social networks in identifying MH disorders, where researchers incorporated public information belonging to other social media users engaged through followings, likes, comments, and tweet replies. For instance, Liaw et al. [134] and Ricard et al. [110] respectively involved liked content and that generated by users who have liked or commented on posts created by individuals of concern. The prior found the amount of depression keywords in liked content to contribute the most performance gain, whereas the latter found an improvement after incorporating such community-generated data. Similarly, Pirayesh et al. [138] and Mihov et al. [139] incorporated content created by homogeneous friends identified through clustering and computation and noticed improvement after increasing the number of homogeneous friends and their respective tweets.

3.3.5. Smartphone and Wearable Sensors

Smartphone sensor data could be utilized to gain an understanding of individuals' mobility (e.g., accelerometer, gyroscope, GPS data), sociability (e.g., call logs, text messaging logs, usage of social applications), and environmental context (e.g., ambient light and sound, wireless WiFi and Bluetooth connections). More personalized insights could be obtained by utilizing location semantics; grouping mobile applications into social, engagement, and entertainment categories; and detecting periodicity and routines to infer individuals' behaviors and routines. Wearable devices further complement smartphone sensor data by offering sleep inferences and physiological signals like heart rate, skin temperature, and calories. Research attempts have uncovered potentially significant indicators of the presence or severity of MH disorders, which we explained in detail in the following paragraphs revolving around three primary aspects, i.e., physical mobility, phone interactions, and sociability.

(1) Physical Mobility Features: Studies have shown that negative MH states and greater depression severity are associated with lower levels of physical activity, demonstrated via fewer footsteps, less exercising [154], being stationary for a greater proportion of time [205], and less motion variability [149], whereas a study on the student population showed an opposite trend for increased physical activity [157]. Movements across locations in terms of distance, location variability, significant locations (deduced through location clusters) [177], and time spent in these places [164] were also valuable. For instance, researchers found greater depression severity or negative MH states associated with less distance variance, less normalized location entropy [154,158], lower number of significant visited places with increased average length of stay [158], and fewer visits to new places [205]. In contrast, Kim et al.'s [162] investigation on adolescents with major depressive disorders (MDD) found that they traveled longer distances than healthy controls. Timing and location semantics could further contribute more detailed insights, such as the discoveries of individuals with negative MH states staying stationary more in the morning but less in the evening [205], those with more severe depression spending more time at home [154,175], and schizophrenia patients visiting more places in the morning [206]. Researchers also acquired sleep information either through inferences from a combination of sensor information relating to physical movement, environment, and phone-locked states or through the APIs of sleep inferences in wearable devices. Sleep patterns and regularity were demonstrated to correlate with depressive symptoms [150,158] where individuals with positive MH states wake up earlier [205], whereas MDD patients showed more irregular sleep (inferred from sleep regularity index) [149].

(2) Phone Interaction Features: Phone usage (i.e., inferred from the frequency and duration of screen unlocks) and application usage were potentially helpful. For instance, several studies [158] found a high frequency of screen unlocks and low average unlock duration for each unlock as potential depressive symptoms. However, while Wang et al. [205] demonstrated the association between negative MH states and lower phone usage, the opposite trend was observed in students and adolescents with depressive symptoms who used smartphones longer [150,162,164]. Researchers also investigated more fine-grained features, such as phone usage at different times of the day, where they found schizophrenic patients exhibiting less phone usage at night but more in the afternoon [206]. Additionally, individuals with MH disorders also showed distinctive application engagement, such as Opoku Asare et al.'s [166] findings that individuals with depressive symptoms used social applications more frequently and for a longer duration. Generally, they also showed more active application engagement in the early hours or midnight compared to healthy controls, who showed diluted engagement patterns throughout the day. Meanwhile, Choudhary et al. [212] revealed that individuals with anxiety exhibited more frequent usage of applications from "passive information consumption apps", "games", and "health and fitness" categories.

(3) Sociability Features: Sociability features, such as the number of incoming/outgoing phone calls and text messages and the duration of phone calls, were also potential indicators of MH disorders [164,175]. For instance, negative MH states are associated with making more phone calls and text messaging [205,222] and reaching out to more new contacts [222]. On the other hand, adult and adolescent populations suffering from MDD were revealed to receive fewer incoming messages [149] and more phone calls [162], respectively. Lastly, ambient environments could also play a role since

individuals with schizophrenia were found to be around louder acoustic environments with human voices [206], whereas those with negative MH states demonstrated a higher tendency to be around fewer conversations [205] than healthy controls.

3.3.6. Demographics and Personalities

In addition, demographics and personalities might play a role in an individual's responses to MH disorders. For instance, several studies [25,109] proved that females have a higher tendency to exhibit depressive symptoms than males. Individuals of different genders may also express varying responses to MH disorders to different extents. For instance, Yazdavar et al. [25] found that females expressed depressive symptoms more prominently on social media, implying their strong self-awareness and willingness to share their encounters to seek support. Meanwhile, a study [219] revealed that age, emotions, and the usage of words related to personal concerns are among the most significant indicators for identifying female samples with potential risks of anorexia nervosa, whereas words relating to biological processes were more indicative for male samples. Clinical experts involved in the study further identified gender as one of the most relevant factors to consider in locating anorexia nervosa. Fine-grained visual elements like formant, eye gaze, facial landmarks, and head pose may also vary across genders with depressive symptoms [70].

On the other hand, existing works [286,287] proved the potential association between MH symptoms and personality traits, where pursuing perfection, ruminant thinking and interpersonal sensitivity could be markers of suicide risk [287], whereas conscientiousness and neuroticism exhibited close relations to depression cues [121]. Researchers have estimated the personality scores of study samples based on textual content, for example, using IBM's Personality Insights https://www.ibm.com/cloud/watson-natural-language-understanding (accessed on 10 December 2023) [57,121] or computing the proportion of words relevant to those in perfection- and ruminant-thinking-related lexicons [187]. Specifically, Chatterjee et al. [188] uncovered that 56% of suicidal samples demonstrated the association between low agreeableness and high neuroticism scores with increased suicide ideation, compared to most healthy controls with high agreeableness and optimism scores. Another study [130] also found that individuals with depressive symptoms generally have higher neuroticism and lower optimism scores.

3.4. Modality Fusion

3.4.1. Feature Transformation to Prepare for Fusion

Some studies further applied transformation on extracted features to prepare for fusion by achieving (1) normalization, (2) dimensionality reduction, and (3) feature alignment. Normalization ensures that numerical features share similar scales and are treated equally by ML models. The most common normalization approaches that we observed are min-max normalization, to scale values between 0 and 1, and z-normalization [288], so that values are zero-mean and unit-variance. Since min-max normalization was claimed to preserve data relationships without reducing outlier effects [56], Cao et al. [187] took this inspiration to represent a subject's age relative to the maximum age among all subjects in the dataset. Meanwhile, dimensionality reduction approaches were widely adopted, such as principal components analysis (PCA), singular value decomposition (SVD), and factor analysis. Lastly, for feature alignment, researchers transformed feature representations of individual modalities to align their dimensions through whitening (ZCA) transform (sparse coded feature representations) [49], global max pooling [77], or discrete Fourier transform (express visual features in the time–frequency domain) [66]. On the other hand, several other studies adopted neural networks to enforce the exact dimensions of multimodal representations. For example, using fully connected (FC) layers with the same units to condense features to a uniform dimension [70,112,179], a multilayer perceptron (MLP) [111], or bidirectional gated recurrent unit (Bi-GRU) [106] to transform representations to specific dimensions and a custom transformer-based architecture that applies linear projection to match various representation sizes [174]. In addition, an FC layer was also used to embed categorical variables to be concatenated with continuous variables [170].

3.4.2. Multimodal Fusion Techniques

Multimodal fusion techniques combine features extracted from different modalities (e.g., audio + visual + textual data) into a single representation for training an ML model. Inspired by an existing work [69], we categorized existing fusion techniques into three main classes, i.e., at the feature, score/decision, and model levels. We hereby emphasize that the current discussion excludes scenarios where fusion is not required if modality-specific features are in independent numerical forms, which ML algorithms could be applied directly.

A feature-level fusion is also known as early fusion, where the features of all modalities are concatenated directly before feeding into an ML model. At the score/decision level, instead of features, researchers combined scores/decisions predicted by individual ML models for each modality, such as probabilities, confidence scores, classification labels, or other prediction outcomes, through operations such as AND, OR, product-rule, sum-rule, and majority voting. Fusing these scores would either produce the final outcome or serve as the input to a secondary ML model. There were also hierarchical score/decision-level fusion approaches that aggregate outputs across multiple layers or stages. For example, in Chiu et al.'s [122] user-level depression classification from social media data, the authors first obtained day-based predictions from post-level outputs weighted based on time intervals. Then, they deduced user-level outcomes based on whether day-based predictions fulfilled predefined criteria.

Unlike feature-level fusion, which concatenates features directly into a single representation, model-level fusion methods utilize an architecture or ML model to learn joint representations that consider the correlation and relationships between feature representations of all modalities. For instance, attention-based architectures (e.g., attention layers, transformers with multi-head attention mechanisms) were adopted to learn shared representations incorporating modality-specific representations with varying extents of contributions based on their significance. Meanwhile, cross-attention mechanisms were employed to consider cross-modality interactions. Shen et al. [20] also proposed using dictionary learning to learn multimodal joint sparse representations, by claiming that such representations are more effective than using features directly as the inputs of ML models. Nevertheless, we acknowledge the limitation that our categorization is merely based on our understanding, and specific fusion techniques in each category may implicitly involve a combination of various fusion levels. For a complete list of studies, methods, and tools, see Appendix B.

3.5. Machine Learning Models

Previous studies adopted ML models for binary classification on the presence of specific MH disorders, multi-class classification on the stages of MH disorders, and regression on the score based on an assessment scale. A complete overview of these models and their application methods is available in Appendix C. Referring to an existing study [289], we classified them into:

- Supervised learning—trained on labeled input–output pairs to learn patterns for mapping unseen inputs to outputs.
- Neural-network-based supervised learning—a subset of supervised learning algorithms that mimics the human brain by having layers of interconnecting neurons that perform high-level reasoning [290] to recognize underlying relationships in data [291].
- Ensemble learning—combines multiple base learners of any kind (e.g., linear, tree-based or NN models) to obtain better predictive performance, assuming that errors of a single base learner will be compensated by the others [292].
- Multi-task learning—attempts to solve multiple tasks simultaneously by taking advantage of the similarities between tasks [289].

- Others—incorporates semi-supervised, unsupervised, or combination of approaches from various categories.

3.5.1. Supervised Learning

The availability of ground-truth information, obtained via expert annotations or clinical assessments, has enabled the broad application of supervised learning approaches that learn the association between input data and their labels. In our findings, these approaches primarily cater to univariate features, where feature engineering may be required to apply them to multidimensional data. The more popular ML algorithms for supervised learning are linear regression, logistic regression, and support vector machines (SVMs). Based on comparisons conducted in existing studies, stochastic gradient descent [43] and least absolute shrinkage and selection operator (lasso) regression [200,213] models performed the best in respective investigations on different feature combinations, i.e., the prior on audio, visual and textual features and the latter on wearable sensor signals, but these models are yet to be compared under similar settings. In addition to the traditional or linear algorithms mentioned above, the following subsection discusses a subset of supervised learning approaches utilizing neural networks.

3.5.2. Neural-Network-Based Supervised Learning

A neural network (NN) [290] is fundamentally made up of an input layer, followed by one or more hidden layers, and an output layer. Each of these layers consists of neurons connected through links associated with weights. An FC layer is included in specific architectures to perform high-level reasoning since it connects all neurons in the previous layer to every neuron in the current layer to generate global semantic information [291]. Meanwhile, an architecture is considered a deep neural network (DNN) when more hidden layers are involved. Although NN architectures could be utilized for various learning approaches, such as supervised, semi-supervised, unsupervised, and reinforcement learning [293], this subsection only concerns those utilized for supervised learning tasks. In such contexts, an NN algorithm approximates a function that maps data received by input neurons to outputs via output neurons by adjusting weights between connected neurons [290]. Therefore, NNs can receive numerical data and yield outputs of any dimension, aligning with the corresponding number of neurons in the input and output layers, respectively.

Throughout this work, we have observed vast applications and versatilities of NN-based models in feature extraction, modality fusion, and ML prediction, which could be applied directly to multidimensional signals or transformed feature representations. As such, we raise the attention of future researchers to the potential overlapping between the NN-based approaches adopted in the three stages above. For example, the outputs of specific hidden layers in such models applied to raw signals or low-level features could be extracted as high-level feature representations, whereas those from the output layers could be utilized as prediction outcomes. The NN model in such scenarios could then be treated as either a feature extraction technique or an algorithm. The same applies to specific sophisticated architectures proposed to capture cross-modality interactions in model-level fusion, where these networks learn fused representations while simultaneously generating predictions.

The abundance incorporation of LSTM [294] for its capability of capturing temporal information across long sequences emphasized its potential. Transformer-based models [295], such as BERT [266] (including its variants like RoBERTa [296], ALBERT [297], EmoBERTa [298]) and XLNet [285], also gained popularity due to their capability to effectively capture contextual information through positional encodings [129] and attention mechanisms to learn different significance weights of relevant information. In contrast, some researchers incorporated attention mechanisms into existing NN architectures such as FC layers, LSTM, and GRU to achieve such emphasis. Despite demonstrating satisfactory efficacy, existing researchers obtained inconsistent findings regarding the influence of NN architecture complexity on the resulting effectiveness. For example, stacking NN architectures, like GRUs [119], CNNs [60], and LSTM [215], improved performance on top of utilizing baseline architectures such as those on both hand-crafted univariate features and raw signals. However, a few studies proved simple shallow NN-based models to succinctly outperform deeper architectures, for instance, AlexNet outperforming VGG-16 and RestNet101 [122], and a 2-layer Bi-LSTM which outperformed LSTM and GRU of varying layers [220].

Overall, the capabilities of NNs in learning high-dimensional data offer promising effectiveness and flexibility in mental healthcare involving heterogeneous data for capturing multifaceted aspects of MH disorders. Nevertheless, such models require large, high-quality datasets since they can only learn patterns within the training data [290]. Due to the complex and non-linear structure with multiple hidden layers, black-box NNs further introduce challenges in obtaining interpretable explanations of how the algorithms arrive at an output [16,299].

3.5.3. Ensemble Learning

Ensemble learning algorithms have shown remarkable effectiveness by combining base models with similar or complementary learning principles [173]. Similar to supervised learning, such algorithms were applied to univariate inputs, which could be hand-crafted numerical features or predicted outputs (e.g., regression scores, probabilities, binary labels) from other baseline models. The few popular ensemble learning approaches are tree-based, such as random forest (RF) [300], eXtreme Gradient Boosting (XGBoost), AdaBoost [301], and Gradient Boosted Regression Tree [302], which utilize decision trees as fundamental. XGBoost and AdaBoost were gradually favored by researchers due to their better predictive performance. Specifically, few studies [134,158,184,212] revealed XGBoost as the most effective among SVM, RF, K-nearest neighbor, logistic regression, and DNN models. In contrast, researchers also proposed novel hierarchical ensemble architectures by stacking algorithms (e.g., XGBoost [194], Extreme Learning Machine (ELM) [192]) into layers where models in subsequent layers receive outputs from previous layers as inputs for ensemble predictions. For example, Mishra et al. [185] and Liu et al. [123] adapted the feature-stacking [303] approach by utilizing logistic regression to combine predictions of various first-level learners, like SVM, KNN, and Lasso regression, applied independently to different feature sets. In addition, Tabassum et al. [168] combined an LSTM-based model applied to hourly time series sensor data and an RF on statistical features aggregated across the data collection duration to benefit from the strengths of respective learning algorithms.

3.5.4. Multi-Task Learning (MTL)

Unlike ensemble learning, MTL involves a single model (of any category mentioned above) trained to solve several simultaneous tasks to exploit task-specific similarities and differences. Examples of task combinations are (1) regression and classification [74,102,118,141,193], (2) depression prediction and emotion recognition [46,106,132], and (3) gender-specific predictions [70]. Though most of the included studies adopted NN-based models, such as CNN [61,62], LSTM [46,106], and DNN architectures [102,193], MTL could also be achieved with linear models, for example, the multi-output support least-squares vector regression (m-SVR) [304] trained to map multivariate inputs to multivariate outputs [207]. Meanwhile, Oureshi et al.'s [70] findings further justified the role of demographics in locating MH disorders, such that incorporating gender prediction as an auxiliary task improved the overall performance.

3.5.5. Others

We observed a few studies applying unsupervised techniques, with clustering using K-nearest neighbors being the most common approach. A few other researchers also adopted anomaly detection using existing unsupervised techniques like Isolation Forest (ISOFOR) [166], or statistical measures such as t-tests for detecting outliers among preliminary prediction outcomes [163]. The research attempts mentioned above revealed that these unsupervised approaches appear more promising on smartphone sensor data than conventional ML approaches, including SVM, RF, GDBT, and MLP. In addition, AbaeiKoupaei et al.'s work [196] was the only semi-supervised learning we identified in this study, in which the authors employed a ladder network classifier [305] consisting of stacked noisy encoder and denoising autoencoder [306]. There were also novel approaches adapting various concepts, including recommender system (RS) [173,307], node classification [173], and federated learning [168]. Additionally, some studies employed computations to learn association parameters [83,189] or deduce prediction outcomes from distance-based homogeneity [85].

3.6. Additional Findings

3.6.1. Modality and Feature Comparisons

Most studies on multimodal detection justified the effectiveness of combining multiple modalities due to their complementary outcomes, which outperformed unimodal approaches. Notably, we noticed a single exception in a finding [174] that textual modality alone is succinctly effective, such that combining it with audio and visual modalities slightly deteriorated the overall performance. Deeper analyses also revealed that specific modalities could be more influential than others. From audio-visual recordings, semantic content in audio transcriptions generated via textual features was found to be more indicative of depression than audio and visual features in several studies on depression, bipolar disorder, and suicidal ideation. Specifically, we found such prominence arising from textual representations using various embedding techniques like GloVe [49], Universal Sentence Encoder [59], Paragraph Vector [102], and ELMo [116].

On the contrary, several revelations highlighted the great potential of audio MFCC features. For example, a study [65] attempting to detect depression in audio samples of less than 10 seconds, another [72] conducted on Chinese language audio samples, and one on detecting bipolar disorders [103] found MFCC features more effective than textual embeddings. Nevertheless, more fine-grained comparisons are required to justify the efficacy of one modality or modality-specific feature over the other due to the varying influence of experimental contexts and setups in data collection and feature extraction.

3.6.2. Personalized Machine Learning Models

In conjunction with an existing finding that individuals with similar depression scores may portray behavioral differences under similar contexts [156], several researchers attempted to achieve individual personalization by training subject-specific models [164,169,205,207], fine-tuning subject-specific layers [161] in a global NN architecture, and deducing personalized predictions by incorporating information from other samples homogeneous to each individual based on correlation coefficients [156] or demographics [208] such as age [209].

Meanwhile, existing attempts at gender-based subgroup personalization also highlighted the potential significance of gender in identifying MH disorders. Researchers achieved such personalizations via training the same ML models on gender-specific samples [92,94,219], fine-tuning and building individual ML models for each gender subgroup [48,161], or incorporating gender prediction as an auxiliary task in an MTL approach [70]. Nevertheless, existing researchers found contradicting findings of models constructed from gender-specific samples. For instance, Pampouchidou et al. [48] and Samareh et al. [54] proved that gender-based classification models outperformed gender-independent ones, whereas others [92,94] demonstrated that global models trained on all genders predicted gender-specific evaluation instances more effectively than those trained on gender-specific data. Attempts above [92,219] further uncovered challenges in effectively predicting female samples, where the outcomes indicated that gender-specific models trained and evaluated on female samples perform worse than those of male samples.

4. Discussion

4.1. Principal Findings

This section addresses the research questions based on findings in the previous section.

4.1.1. RQ1—Which Sources of Data Are Most Effective for Supporting the Detection of MH Disorders?

Our findings have explored evidence of associations between specific modality-specific features and various MH symptoms. In Figure 5, we categorized MH symptoms into psychological, physical, social, and physiological aspects [19] and mapped these aspects to the capabilities of data sources in capturing them. The figure illustrated that no multimodal data source can capture all aspects of MH symptoms and that deducing the most effective data source relies on the symptoms researchers wish to investigate that are relevant to specific MH disorders (see Table 4 for existing studies utilizing specific data source to investigate each MH disorder). We acknowledge that the figure only includes mapping relevant to passive data sources included in the current study and that other active sensing approaches may be valuable and complementary.

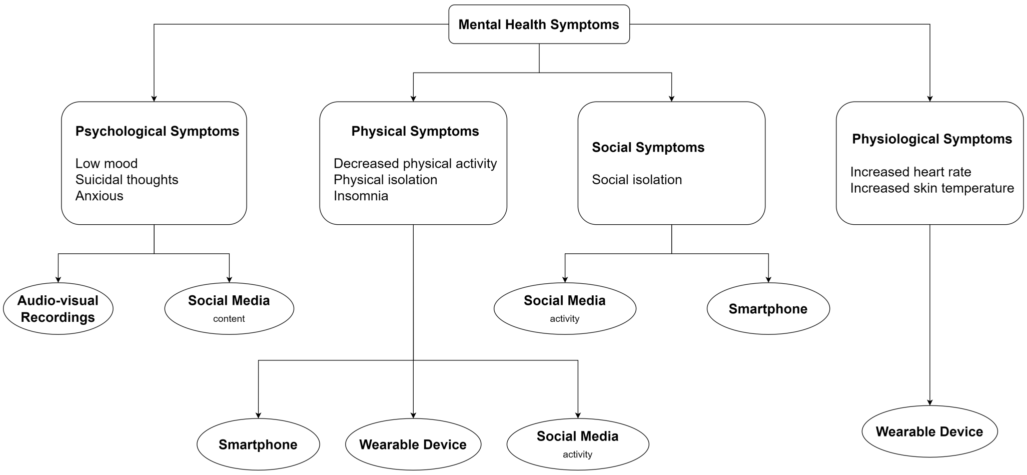

Figure 5. Mapping of data source to mental health symptoms.

Psychological symptoms related to moods, emotions, and feelings were shown to be effectively captured by textual features, which could be obtained from transcriptions of audio-visual recordings and the content of social media posts. For example, individuals with MH disorders expressed stronger negative emotions via texts with overall negative sentiment or more negative words and emoticons, as well as through MH-specific keywords related to symptoms, treatment, and medications (e.g., antidepressant names and phrases associated with depressive symptoms for depression). While these textual features have been proven indicative, researchers found visual cues to provide complementary information by encapsulating finer details of individuals' implicit inner emotions. For example, the significant association between FAUs and depressive symptoms may indicate how individuals present their facial expressions in response to MH symptoms. In addition, individuals' publicly shared images may reflect their psychological conditions; for instance, preferences for darker colors may represent lower moods, or images of animals may represent self-coping mechanisms to improve emotional states.

Meanwhile, for physical symptoms, the unobtrusiveness and ubiquity of smartphones and wearable devices have great potential to capture individuals' natural behaviors, which could reflect the physical manifestations of MH-specific symptoms. For example, the association of higher depression severity with lower physical mobility, demonstrated via being stationary for a greater proportion of time or traveling to fewer places sensed using GPS and accelerometer, may suggest depressive symptoms of losing interest in surroundings, lethargy, or social isolation. Wearable sensors could complement by offering sleep-related information like sleep states and duration to infer sleep quality. In addition, individuals' social media activities could reflect their personal routines and behaviors, for example, higher susceptibility to insomnia or sleeping problems implied through frequent posting activities during midnight. In contrast, a decline in social interactions is an example of social symptoms that might indicate a reduction of interest in surroundings. Such social interactions could include both verbal communications detected via microphones in smartphones and social interactions made through social media platforms and social mobile applications. Lastly, wearable devices are the only passive data source capable of tracking changes in physiological symptoms, including heart rate, skin temperature, and calories burnt.

Nevertheless, our findings in Section 3.3.6 highlighted the influence of demographics and personalities on individuals' behaviors, where specific subgroups may openly share their symptoms to seek external support [25], while some may show reluctance through prominent usage of the word "others" or third-person pronouns on social media [123,187]. Additionally, Shen et al. [109] proved the divergence in Twitter and Sina Microblog user behaviors, where those with depressive behaviors posted less frequently on Twitter, but the opposite trend was observed in Sina Microblog users. There was also a higher occurrence of positive words in textual content in the latter than in the prior. Such disparity in expressivity could be attributed to different populations' cultural and language differences since Sina Microblog users are primarily from Asian countries. Meanwhile, an increased regularity in physical activities could be a coping mechanism for young adults [157] but is not necessarily the case for other contexts or populations having a depressive symptom of lacking interest in activities, as shown in several other studies [110,149,150,152,158].

In addition to modality-specific effectiveness, discussed in Section 3.3, there is a need for deeper considerations beyond data source, which are associated with the experimental contexts of data collection approaches and the individuals to which the data belong. We established several criteria to evaluate data collection approaches in greater detail in Section 4.2 below.

4.1.2. RQ2—Which Data Fusion Approaches Are Most Effective for Combining Data Features of Varying Modalities to Prepare for Training ML Models to Detect MH Disorders?

Based on our categorization of modality fusion techniques, namely feature, score/decision, and model levels, we recommend employing feature or model-level fusion based on researchers' specific use cases. Our observations suggest that score-level fusion might be less effective, as modality features are modeled separately by individual ML algorithms, thereby ignoring the potential correlation of features across different modalities. As previously discussed, certain modalities may be more effective than others in unimodal settings. Since score-level fusion only considers the intermediate prediction outcomes of modality-specific ML models, the more effective modalities may overshadow the less significant ones in the final outcomes, even though all modalities are complementary. The following paragraphs provide more detailed recommendations to decide between feature and model-level fusion.

Feature-level fusion is readily applicable to simple univariate features, where direct concatenation is straightforward and efficient to implement. Researchers should align their decision with research objectives by considering whether the features succinctly capture the information they intend to investigate with sufficient details and the capability of adopted ML algorithms to model the correlation among such features to answer specific research questions. For example, a high-level aggregation by computing the average steps, distance traveled, and time spent at an individual's home across the entire study duration produces a simple univariate vector representation for feature-level fusion, but such information may not be relevant for a study intending to capture daily variations in individuals' physical behaviors. If researchers intend to prioritize efficiency and low computational complexity, specific computation methods should be investigated to effectively elicit low-dimensional representations that capture time-based variations. For instance, autocorrelation analysis captures feature periodicities across specific durations [157] where the resulting correlation coefficients could be utilized as higher-level representations of time series data.

On the other hand, model-level fusion has been shown in several attempts [69,71,73,308] to generate more effective high-level feature representations than hand-crafted univariate features due to their capability of capturing temporal and contextual dependencies while modeling cross-modality interactions. The decisions of whether to adopt model-level fusion and the architecture to employ should consider the complexity of the research problems. For example, deep NN architectures with more extensive layers may be more effective if researchers are interested in investigating the influence of particular factors across long durations. However, researchers should be aware of the greater computational costs associated with more complex architectures and the black-box nature of certain NN-based architectures, which reduce the interpretability of modeled interactions.

We have observed an increased adoption of various NN architectures among researchers to model feature information at varying complexity levels. Several architectures have shown outstanding efficacy by incorporating both temporal and contextual interactions within and across modalities, underscoring the importance of generating fused representations that encapsulate such information. For instance, Yan et al. [170] applied Convolutional Sequence Embedding Recommendation (CASER) [309], which leverages convolutional filters of CNNs. In CASER's horizontal convolutional layer, the authors applied convolutional filters horizontally to capture daily-level sequential patterns as local features for all feature points at the previous time step, followed by max-pooling to extract the most meaningful information. In the vertical convolutional layer, feature-level patterns were generated as the weighted sum of each feature point at specific time steps, with the convolutional filter acting as weights. The outputs of both convolutional layers were then concatenated into fully-connected layers to produce fused representations. This architecture demonstrated more effective capture of hidden series patterns than aggregated statistical features (e.g., average, minimum, or maximum across a duration). In contrast, Zhou et al. [93] proposed a time-aware attention multimodal fusion (TAMF) network. This architecture includes a sparse MLP, utilizing its weight sharing and sparse connections to mix information from modality-specific representations in both vertical and horizontal directions. The resulting outcome is a mixed attention vector, which is separated into attention vectors of each modality. The final fused representations were obtained by summing modality-specific representations weighted by respective attention vectors. TAMF was claimed to model the importance of different modalities at different times, with well-rounded consideration of cross-modality interactions.

4.1.3. RQ3—What ML Approaches Have Previous Researchers Used to Successfully Detect MH Disorders from Multimodal Data?

We could not deduce a single one-size-fits-all model that is the most effective for various multimodal tasks. This is because the effectiveness of ML algorithms relies on the nature and structure of the data, the task to achieve, and how data information is learned and fully utilized. Nonetheless, our observations revealed that ML models adopted in existing studies are primarily supervised learning algorithms, which utilize ground truth as the "gold standard", and several models worth investigating are Lasso regression, XGBoost, LSTM-based models, and transformer-based models like XLNet, based on their prominent predictive performance in existing studies. Notably, we noticed a similar trend of rising adoption of NN-based models in recent studies, which appear more valuable than linear statistical algorithms in modeling multidimensional time series signals. As previously discussed in Section 3.5.2, existing researchers have proposed various novel architectures to incorporate temporal, contextual, and cross-modality dependencies, such as injecting time-based representations into transformer models to improve performance further [129]. We also observed the significance of utilizing relevant data information, where a few studies [111,186] demonstrated that irrelevant textual and visual content introduced noise that obstructs traces of MH disorders and caused further performance deterioration. While the studies above selected more relevant information through techniques like reinforcement learning [111], other studies utilized attention mechanisms to exert significance weights based on relevance instead of eradicating less relevant information. Specifically, attempts [128,146] at hierarchical attention asserted onto deep representations of social media content from word level and subsequently to post and user levels highlighted the great prospect of such mechanisms.

Despite both audio-visual and sensor data being time series data, we noticed relatively limited applications of NN architectures on sensor data. Most studies employed conventional linear or statistical ML algorithms (e.g., logistic regression, SVM, XGBoost), which learn from univariate inputs, by aggregating extracted features across the whole duration. Such approaches potentially neglected the associations of features across time since several recent studies [162,170,176] proved the superiority of NN-based models applied to higher-dimensional time series sensor data, where features are aggregated at hourly, daily, or weekly features, over conventional univariate approaches. These outcomes suggested an aspect worth investigating for future researchers to better harness the potential of ML algorithms, for example, by applying an LSTM-based model to hourly time series data and RF to univariate features derived from the prior data to leverage the strengths of both algorithms [208].

While most studies predominantly focused on optimizing performance metrics like accuracy, precision, and recall, they often overlooked practical considerations for real-world applications of ML models, such as complexity, explainability, and generalizability. Despite the inherent biases in ML algorithms [310], only two included studies examined potential biases at the individual and gender levels. For ML models to be seamlessly integrated into real-time detection applications for clinical use, they must be lightweight in terms of complexity and computational cost, considering that available memory and computation resources may be restricted [311], especially in individuals' local devices. Specifically, in the MH domain, this criterion is significant to enhance efficient computations on the fly and timely delivery of personalized interventions and recommendations without imposing excessive processing power [311]. Ultimately, achieving this can potentially improve individuals' accessibility to mental healthcare resources and subsequently promote their treatment-seeking.

In addition, with the growing attention to the interpretability and explainability of ML models [299], these criteria offer transparency to ML models' decision-making to establish trust in these algorithms and elevate their practicality in real-life applications. Specifically, in high-stake MH applications where black-box predictions potentially bring harmful consequences, explanations of ML outputs can provide meaningful insights for healthcare professionals to understand and validate the relevance of ML outputs in complementing clinical diagnosis. Considering the inherent biases in ML algorithms [310], transparency in working mechanisms (local explainability), feature contributions (global explainability [299]), and potential shortcomings are necessary for healthcare professionals to guide and manage the influence of ML outputs on clinical decisions.

Meanwhile, generalizability improves the transferability of ML models to external scenarios beyond local training environments. Given the potential influence of demographics and personalities on manifestations of MH-related behaviors, existing surveys [17] have highlighted the risks of under-representation of certain groups in training datasets, in which the demographic disparities may be magnified in the subsequent applications to the MH domain. Generalizability can enhance the applicability of ML models to other contexts and heterogeneous populations, subsequently improving the accessibility of the general population to MH resources. As an existing work [312] demonstrated the difficulties of aligning cross-study settings for improved generalizability, Thieme et al. [17] emphasized avoiding overclaiming premature generalization from datasets lacking clinical validation and diversity. As such, future researchers should validate and communicate potential limitations in the generalizability of research outcomes. For example, researchers should account for the diversity within a population by validating outcomes across different subgroups with various demographics and characteristics and reporting on the metadata of the community from which the data is collected.

4.2. Evaluation of Data Sources

Following the address of RQ1 in Section 4.1.1 above, we established several criteria to further analyze the different categories of data sources.

4.2.1. Criterion 1—Reliability of Data

The reliability of a data source relies on how well it captures people's real-life behaviors. This criterion is crucial to contribute relevant data for supporting clinical diagnosis since failure to reflect realistic behaviors may result in misdiagnosis of MH disorders, potentially leading to severe complications.

We perceived that smartphone and wearable sensor data are the most reliable due to their ubiquity and a lower possibility of people "tricking" sensors into gathering perceivably desired data. Prior to data collection, participants will configure dedicated mobile applications in their smartphones, allow permission to access specific sensor data, and establish wireless connections for wearable devices to their smartphones, where applicable. Afterwards, they will interact with their devices as usual throughout the data collection process with minimal active inputs. Given that they are open to and allow monitoring over a longer period, the awareness of monitoring may be reduced following the initial novelty effect, thereby enhancing the "honesty" of corresponding data. Nevertheless, researchers should consider data quality that can be affected by sensors of different devices with varying sensitivity. The fit of wearable devices may also affect the accuracy and amount of data collected. For example, improper wearing or wearables slipping off [313] during sleep may end up collecting poor, noisy data. In addition, participants may forget to reapply sensors (e.g., following a shower) or feel discomfort from wearing it on their wrists or other parts of their body.

Both audio-visual recordings and social media data potentially suffer from biases introduced by individuals' self-presentation concerns. For example, a person may behave differently under the pressure of continuous supervision, also known as the Hawthorne effect [314], to look generally appealing or hide any indicative behaviors potentially due to fear of judgment. Similarly, social media users might curate their public posts to appear presentable due to factors like the consciousness of unfavorable public perception or social stigma. In addition, a study [187] found that users express their thoughts differently in the hidden tree hole posts than in usual Sina Microblog posts. A tree hole is a microblog space whose author has committed suicide, and other users tend to comment under the last post of the passed one about their inner feelings and thoughts. Such posts were revealed to contain more self-concern and suicide-related words, thereby challenging the detection of MH through regular or public posts.

4.2.2. Criterion 2—Validity of Ground Truth Acquisition

A well-justified ground truth is vital to represent people's actual MH states. From a responsible innovation perspective, unrealistic ground truth can cause under or overestimations in MH detection, which may escalate to introduce dangers, especially in disorders with crisis points, such as suicidal ideation or an eating disorder.

Clinical assessments are the only method yielding representative ground truth thus far because self-reports, whether in the form of responses in assessment scales or self-declaration in social media posts, are subject to self-presentation and recall biases. However, we noticed a possibility of verifying ground truth from self-reports by utilizing behaviors detected via smartphone and wearable sensing. These approaches typically request individuals to answer clinically validated assessment scales based on guidelines like the Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition (DSM-V) [315] to serve as baseline ground truth. The response to each assessment question corresponds to recalled behaviors over a specific duration. Taking Wang et al.'s work [150] as inspiration, behaviors inferred from sensor data can be mapped to individual DSM-V symptoms to verify ground-truth labels. Conversely, some researchers primarily rely on social media users' self-identification of MH disorder diagnosis to acquire the ground truth of their

MH states, usually through identifying keywords associated with specific MH disorders. Such acquisition risks under-identification since it depends on whether people took the initiative and felt comfortable sharing the information. Solans Noguero et al. [219] further proved that suicide-related lexicons were less comprehensive due to the likelihood of omitting explicit vocabulary and failing to identify implicit hints.

A reliable ground truth should always be supported by clinical validation, such as through a diagnosis by trained practitioners, reference of clinical evidence, or having clinical experts verify manual annotations, since relevant clinical knowledge is necessary to ensure the validity of ground-truth labels. Specifically, in the use case of social media data, where individuals' data were crawled directly from these platforms, there are both practical and ethical considerations that need to be addressed when claiming a ground truth has been established. While time-consuming, future studies could consider ways to directly approach social media users where possible to verify their MH states and to actively gain their consent for their data to be used (or ensure that users are aware of the research aims at the very least).

4.2.3. Criterion 3—Cost

We inspected costs in terms of (1) data accessibility, (2) external costs incurred for dedicated data collection equipment and tools, (3) processing power for transforming and analyzing data, and (4) storage space. These considerations are crucial in evaluating the practicality of research outcomes in real-life applications so that a cost-effective method can be easily deployed to benefit the target population.

We deduced that social media data are the cheapest to acquire from all aspects above. It is the most accessible since researchers can crawl public data online without accessing users individually or getting hold of their private information, given that they comply with the platforms' terms and conditions (whether this is deemed ethical is another question). Relatively small processing power and storage space are required since crawled data is in the form of data entries. Additionally, features can be extracted from textual and visual content using processing tools available, such as LIWC [278], NLTK [280], and SEANCE [282] for texts and OpenFace [269], OpenCV [270], and OpenPose [271] for images. Audio-visual recordings are the most costly because they encapsulate rich data information that requires large storage space and extensive computation power to process audio and visual elements. In addition, this approach requires video cameras and microphones, which might have to be purchased beforehand, and consumes more effort in setting up the equipment at one or more locations based on device reception and coverage.

Since most populations generally own smartphones [316], the potential equipment cost for smartphone sensing is lowered. However, a substantial cost might be incurred if researchers are to provide smartphones to study participants without smartphones or to ensure consistency. In contrast, wearable devices are cheaper but less ubiquitous than smartphones since some participants do not see the necessity of possessing wearable devices (e.g., fitness watch, smartwatch) and consider them a luxury item. Though both approaches involve time series sensor signals, which may be high dimensional, the storage cost is still relatively economical compared to multidimensional video files. Nevertheless, we observed a novel application of federated learning in Tabassum et al.'s [168] work, which is potentially feasible for resolving storage and privacy concerns. The authors processed collected data and extracted features locally in individuals' devices to obtain local features and parameters. These were then utilized to fine-tune local individual-specific ML models, which share and exchange higher-level parameters with a global-centric model. This approach significantly lowered transmission cost and storage space since server transmission is reduced from complex multidimensional data to numerical features and parameters while minimizing the risks of privacy leaks during transmission since raw data was discarded after local processing. As such, researchers should consider the factors discussed in Section 4.1.3 to ensure that ML models residing in local devices are highly deployable, such as being efficient and lightweight, to avoid consuming excessive local processing power.

4.2.4. Criterion 4—General Acceptance

The general acceptability of people towards specific data collection approaches has the most direct influence on research involving human data. This criterion can be attributed to people's openness and comfortability in allowing their data to be collected, which are often supported by their perceptions and concerns about the methods. The control they have over the sharing of their data may also be a contributing factor.

We inferred wearable sensing as the most acceptable because it gathers the least identifiable data (e.g., physiological signals like heart rate and skin temperature, activity levels, and sleep patterns) that is most unlikely to disclose people's personal information. On the other hand, acceptability towards smartphone sensing is debatable. A study [100] discovered GPS to be the most acceptable compared to calendars, call logs, text logs, and contacts, and only one-third of study participants shared their smartphone logs. However, this is not necessarily the case for some with safety concerns about revealing their locations (e.g., not wanting to disclose their homes or concerns over being stalked [317]), and allowing access to call and text logs can also be perceived as privacy-invasive. In both smartphone and wearable sensing contexts, participants may or may not have control over the kinds of data collected from them, i.e., which sensors are enabled, depending on the approach design and configuration by researchers.

The acceptance of social media users for researchers to utilize their data for research purposes is also controversial. Researchers have presumed that social media users are open to and permit others to access their data since they opted to make it public in the first place [318]. Even though users have complete control of their public content, they are unaware and may oppose their data being accessed and analyzed for research without consent. Meanwhile, we hypothesize that audio-visual recordings are the least acceptable because it is highly invasive, and not all individuals are comfortable having their footage taken and monitored continuously. Even though existing research [313] found a general acceptance of being recorded using privacy-preserving video cameras that only capture participants' silhouettes, such cameras may not apply to the current context that requires identifiable elements, like facial expressions, body gestures, and movements. Researchers have complete control of the data collection process, and there were contradictory opinions from participants themselves on whether they should have control over when and what is being recorded, e.g., by allowing them to pause at specific critical times [313].

As much as the ability to manage data sharing based on personal comfort can improve the acceptability of data collection, researchers should be aware of the resulting risks of biases and sparsity in data. A reduction in the unobtrusiveness of passive sensing and an increased likelihood of skewness will occur if participants constantly manage their data sharing. There will also be data sparsity issues if participants can selectively activate/deactivate specific sensors at random times. There are other means of establishing people's trust in researchers to raise their confidence that their shared data will be kept secure and handled cautiously with safety procedures. For example, this can be achieved by offering transparency of what is being collected, why, and how they will be stored and handled.

4.2.5. Overall Findings

Overall, smartphone sensing emerged as the most promising avenue. Our findings demonstrate abundant significant correlations between sensor features and MH symptoms, offering the potential to translate such connections into an individual's physical manifestations in response to specific MH disorders. While symptoms associated with specific MH disorders may manifest differently, the capability of smartphone sensors to capture natural behaviors and variations across time provides a strong advantage.

However, the integration of smartphone sensing into MH applications demands further research due to several critical considerations that are yet to be addressed. Ethical concerns arise regarding whether it is privacy-infringing for researchers to access individuals' private or personal behaviors, which they may be unwilling to disclose. Consequently, they may deliberately hide or alter their behaviors to "trick" the data collection system or withdraw due to privacy concerns. These issues contribute to the potential unreliability and sparsity of the resulting data, introducing challenges for technical researchers to seek solutions for data-driven ML algorithms, especially in supervised learning with a heavy reliance on ground truth labels. Despite the rich information in time series sensor-based data, there is still room for research to investigate techniques that fully harness its potential. While existing studies have demonstrated the efficacy of neural network architectures in modeling such high-dimensional data, the low explainability and high complexity of such architectures remain a critical challenge. Additionally, existing elicitation techniques are often informed by standard guidelines like DSM-V, potentially disregarding behaviors yet to be discovered. As such, it is imperative to establish a common ground between researchers and clinical experts that enables collaboration to investigate and interpret ML outputs to ensure clinically relevant outcomes.

We hereby acknowledge that the above represents our perspectives based on the current understanding and analysis and that it is essential for future researchers to critically evaluate and adapt the insights based on the evolving landscape of technology and methodological approaches to their specific use cases in the MH domain. We outline some guidelines in the following subsection to assist future researchers in making informed decisions.

4.3. Guidelines for Data Source Selection

In light of the various influencing factors of MH conditions and the necessary considerations for high-stakes applications involving vulnerable individuals, we have devised guidelines that future researchers can use in conjunction with Figure 5 above for selecting an optimal data source or combinations of data sources based on specific use cases.

1. *Define research objectives and scope:* Clearly defined research objectives and questions can guide researchers to determine the kind of information required to achieve the research goals and, subsequently, to evaluate the extent of the data source in accurately representing or capturing relevant information. Determining the scope of the study is crucial to pinpoint and assess the relevance of data information to ensure that collected data effectively contributes to the desired outcomes.
2. *Determine the target population:* Identifying the target population and its characteristics involves various aspects, including the targeted MH disorders, demographics, cultural backgrounds, and geographical distribution. These aspects are mutually influential since individuals' behaviors and data may vary based on reactions to different MH disorders, with further influence caused by cultural backgrounds and demographics, such as age, gender, and occupation. Additionally, geographical distribution and economic backgrounds may influence an individual's accessibility to a specific data collection tool. This consideration ensures that the data collected is representative and applicable to the population of interest, enhancing the overall effectiveness of the approach.
3. *Identify candidate data sources and evaluate their feasibility:* Evaluating the feasibility of each data source in light of the research objectives and target population identified above assists researchers in making informed decisions. Given the contexts and environments in which the target population is situated, researchers can assess which data source is the most practical and relevant. For example, researchers may consider employing remote sensing to introduce the unobtrusiveness of data collection for high-risk MH disorders or overcome geographical challenges. This assessment should consider its feasibility in terms of cost and accessibility, and it should be informed

by Figure 5 to ensure that the selected data source can effectively capture relevant MH symptoms.

4. *Consult stakeholders:* Engaging stakeholders, including healthcare professionals, patients, and families, provides various perspectives of parties involved in supporting individuals with MH disorders. These consultations verify and offer insights into the acceptability and feasibility of data sources and help ensure that researchers' decisions align with ethical considerations and stakeholders' comfort.
5. *Ethical considerations and guidelines:* Researchers should further consult institutional review boards and established guidelines to ensure the compliance of data collection procedures with ethical standards and research practices. This step is crucial to safeguard participants' rights and privacy, enhancing the credibility of the study.
6. *Assess the significance of ground truth information:* Evaluating the significance of ground truth information informs how researchers gauge its impact on the study and whether specific workarounds are necessary to enhance ground truth reliability and validity during data collection. This evaluation will then aid researchers in designing the data collection procedure and determining the extent of reliance on ground truth to support future analysis, reasoning, and deductions.

5. Conclusions

This study examines existing methodologies for non-intrusive multimodal detection of MH disorders and critically evaluates various data sources in terms of reliability, ground truth validity, cost, and general acceptance. Given the complexity of identifying the most effective data source for detecting MH disorders, our guidelines offer a systematic approach for future researchers to make informed decisions about a data source that aligns with research objectives, is relevant to the target population, and adheres to ethical standards. In addition, our analysis highlights the potential of neural network architecture in model-level fusion for capturing higher-complexity cross-modality interactions. We also observe the prospect of utilizing such architectures as ML algorithms to handle high-dimensional data, though practical aspects, such as complexity, explainability, and generalizability, should be scrutinized beyond effectiveness.

We acknowledge the inherent limitations in our approach, recognizing that our search strategy might have omitted potential data sources not explicitly defined within our predetermined categories. Though our findings verified the significance of multimodality compared to unimodality in most cases, there is no absolute answer since the overall efficacy depends on modality-specific features. In addition, there are risks associated with our assumption that passive sensing captures natural behaviors and is more acceptable. The deliberate exclusion of active sensing based on this assumption limits our understanding of potential insights that active sensing approaches can offer. In conjunction with our previous discussion on seeking validation in ground truth information, active inputs may be valuable and necessary to achieve robust validation. As we critically evaluated each data source, we observed a refutation of our assumption, such that passive sensing approaches can be privacy-invasive and are not necessarily well accepted. This is due to the uncertainties and unobtrusiveness of such approaches, which may introduce a sense of insecurity among individuals from whom the data is collected. Building upon the acknowledgements, the current study has recognized smartphone sensing as a promising avenue for further exploration as our next step forward. In light of the ethical considerations and limitations identified, we plan to conduct interviews and focus groups with individuals with MH disorders to gather feedback on the acceptability of smartphone sensing and potential workarounds for addressing privacy concerns. Simultaneously, consulting healthcare professionals will provide valuable perspectives on incorporating smartphone sensing into clinical practice. As we embark on the journey into smartphone sensing, we extend an open invitation for collaboration with fellow researchers, healthcare professionals, and stakeholders passionate about advancing in this domain.

Nevertheless, our work aspires to bring significant implications for stakeholders, including researchers, mental healthcare professionals, and individuals with MH disorders. Our overview of current methodologies for handling multimodal data serves as a starting point for future MH researchers to explore methodological advancements for more effective and timely detection approaches. Our guidelines for data source selection provide a systematic approach for researchers to make informed decisions aligned with use cases or specific symptoms of interest. In addition, our critical analysis of passive multimodal data sources and modality-specific features provides insights to explore the effectiveness of other modality combinations for specific MH disorders. Subsequently, this inspires the development of specific tools that leverage external or multiple data sources to support mental healthcare professionals in their clinical practice (e.g., drawing inspiration from the beHEALTHIER platform [319] which integrates different types of healthcare data, including health, social care, and clinical signs, to construct effective health policies). We envision engaging with MH professionals through workshops, webinars, or other collaborative efforts to bridge the gap between research and practice. Additionally, our practical insights emphasize implementing ML approaches in real-world settings, paving the way for practical implementations that enhance the accessibility for individuals with MH disorders. The outcomes related to the correlation between specific inferred behaviors and MH symptoms also contribute to a better understanding of MH symptoms. Moving forward, we anticipate close collaboration with mental healthcare professionals and individuals with specific MH disorders to design a multimodal approach that facilitates more effective detection. Regardless, we acknowledge the need to establish a middle ground to effectively communicate technical concepts and implications to both stakeholder groups.

Author Contributions: Conceptualization, L.S.K., M.K.L., C.Y.C. and R.M.; writing—original draft preparation, L.S.K.; writing—review and editing, L.S.K., M.K.L., C.Y.C. and R.M.; supervision, M.K.L., C.Y.C. and R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was carried out within the framework of the research project FRGS/1/2020/ICT02/MUSM/03/5 under the Fundamental Research Grant Scheme provided by the Ministry of Higher Education, Malaysia.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AdaBoost	Adaptive Boosting
ADHD	Attention Deficit Hyperactivity Disorder
BDI	Beck Depression Inventory
CES-D	Center for Epidemiological Studies Depression Scale
CNN	Convolutional neural network
DNN	Deep neural network
DSM-V	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ED	Eating disorder
GAD-7	General Anxiety Disorder-7
GPS	Global Positioning System
GRU	Gated recurrent unit
HDRS	Hamilton Depression Rating Scale

LSTM	Long short-term memory
MDD	Major depressive disorder
MFCC	Mel frequency cepstral coefficients
MH	Mental health
ML	Machine learning
MLP	Multi-layer perceptron
MRI	Magnetic Resonance Imaging
MTL	Multi-task learning
NN	Neural network
OCD	Obsessive-compulsive disorder
PHQ-9	Patient Health Questionnaire-9
PTSD	Post-traumatic stress disorder
RF	Random forest
SLR	Systematic literature review
SVM	Support vector machine
XGBoost	Xtreme Gradient Boosting

Appendix A. Existing Modality Features

Table A1. Audio features.

Features	Tools	Studies	Feature Category
Low-level descriptors: jitter, shimmer, amplitude, pitch perturbation quotients, Mel-frequency cepstral coefficients (MFCCs), Teager-energy cepstrum coefficients (TECCs) [320], Discrete Cosine Transform (DCT) coefficients	OpenSmile [267], COVAREP [321], YAAFE [322], Praat [323], Python libraries (pyAudioAnalysis [324], DisVoice [325]), My-Voice Analysis [326], Surfboard [327], librosa [328]	[12,48,51,72,74,78,81,87,88,90–92,94,97,99,101,104,107,108,184,192,195–199,211,214]	Voice
Existing acoustic feature sets: Interspeech 2010 Paralinguistics [329], Interspeech 2013 ComParE [330], extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) [331])	OpenSmile [267]	[51,57,59,61,63,74,81,97,103,192,194–198]	Voice
Speech, pause, laughter, utterances, articulation, phonation, intent expressivity	Praat [323], DeepSpeech [332]	[12,48,79,107,184,193,203,211]	Speech
Vocal tract physiology features	N/A	[49]	Speech
Embeddings of audio samples	VGG-16 [261], VGGish [333], DeepSpeech [332], DenseNet [334], SoundNet [259], SincNet [335], Wav2Vec [336], sentence embedding model [337], HuBERT [338], convolutional neural network (CNN), bidirectional LSTM (BiLSTM), ResNet [308], graph temporal convolution neural network (GTCN) [339]	[59,60,67,72,74,80,86,89,93,96,100,136,174,195]	Representations
Graph features: average degree, clustering coefficient and shortest path, density, transitivity, diameter, local and global efficiency	Visibility graph (two data points visible to each other are connected with an edge)	[81]	Representations

Table A1. *Cont.*

Features	Tools	Studies	Feature Category
Statistical descriptors of voice/speech features: mean, standard deviation, variance, extreme values, kurtosis, 1st and 99th percentiles, skewness, quartiles, interquartile range, range, total, duration rate, occurrences, coefficient of variation (CV)	Manual computation, histograms, DeepSpeech [332]	[12,55,56,91,92,99,107, 193,197,214]	Derived
Bag-of-AudioWords (BoAW) representations of voice/speech features	openXBOW [340]	[59,74]	Representations, Derived
High-level representations of features/representations (capture spatial and temporal information)	Gated recurrent unit (GRU) [341], LSTM, BiLSTM, combination of CNN residual and LSTM-based encoder–decoder networks [75], time-distributed CNN (T-CNN), multi-scale temporal dilated convolution (MS-TDConv) blocks, denoising autoencoder	[61,65,67,73,75,77,87,94, 100,199]	Representations, Derived
Session-level representations from segment-level features/representations	Simple concatenation, Fisher vector encoding, Gaussian Mixture Model (GMM)	[192,199,214]	Representations, Derived
Facial/body appearance, landmarks, eye gaze, head pose	OpenFace [269], OpenCV [270], Viola Jones' face detector [342], CascadeObjectDetector function in MATLAB's vision toolbox, Haar classifier [270], Gauss–Newton Deformable Part Model (GN-DPM) [343], OpenPose [271], ZFace [344], CNN [46], Faster-RCNN (Region CNN) [147], multilevel convolutional coarse-to-fine network cascade [345], Inception-ResNet-V2 [346], VGG-Face [68], DenseNet [334], Affectiva https://go.affectiva.com/affdex-for-market-research (accessed on 10 December 2023), DBFace https://github.com/dlunion/DBFace (accessed on 10 December 2023), FaceMesh https://developers.google.com/android/reference/com/google/mlkit/vision/facemesh/FaceMesh (accessed on 10 December 2023), dlib [347]	[25,53,55,56,68,79,80,83, 91,92,94,98,99,101,108, 174,193,199,201,203,220]	Subject/Object
Appearance coefficients of facial image and shape	Active Orientation Model (AOM) [348]	[50]	Subject/Object
Probability distribution of 365 common scenes	Places365-CNN [349]	[220]	Subject/Object

Table A2. Visual features.

Features	Tools	Studies	Feature Category
Feature descriptors: local binary patterns, Edge Orientation Histogram, Local Phase Quantization, Histogram of Oriented Gradients (HOG)	OpenFace [269]	[47,48,53,195]	Subject/Object, Derived
Geometric features: displacement, mean shape of chosen points, difference between coordinates of specific landmarks, Euclidean distance, angle between landmarks, angular orientation	Manual computation, subject-specific active appearance model (AMM), AFAR toolbox [350]	[47,51,54,56,70,79,83,91, 99,195,198,203]	Subject/Object, Derived
Motion features: movement across video frames, range and speed of displacements (facial landmarks, eye gaze direction, eye open and close, head pose, upper body points)	3D convolutional layers on persons detected at frame-level, Motion history histogram (MHH) [351], feature dynamic history histogram (FDHH), residual network-based dynamic feature descriptor [75]	[52,53,68,75,147,193]	Subject/Object, Derived
Facial action units (FAUs), facial expressions	OpenFace [269], Face++ [352], FACET software [353], AU detection module of AFAR [350]	[25,79,91,99,194,196,198]	Subject/Object, Emotion-related
FAU features: occurrences, intensities, facial expressivity, peak expressivity, behavioral entropy	MHH, Modulation spectrum (MS), Fast Fourier transform (FFT)	[83,99,101,107,194,354]	Emotion-related, Derived
Emotion profiles (EPs)	SVM-based EP detector [355]	[101]	Emotion-related
Sentiment score	ResNeXt [356]	[186]	Emotion-related
Turbulence features capturing sudden erratic changes in behaviors	N/A	[192]	Derived
Deep visual representations from images or video frames	VGG-16 [261], VGG-Face [357], VGGNet [261], AlexNet [358], ResNet [308] ResNet-50 [359], ResNeXt [356], EfficientNet [360], InceptionResNetV2 [346], CNN, dense201 [195], self-supervised DINO (self-distillation with no labels) [361], GTCN [339], unsupervised Convolutional Auto-Encoder (CAE) (replaces autoencoder's fully connected layer with CNN) [195]	[53,58,60,74,82,84,85,89, 93,95,98,106,111,115–117, 122,125,126,129,131,132, 135,160,187,195,201,220]	Representations
High-level (frame-level) representations of low-level features (LLDs, facial landmarks, FAUs)	Stacked Denoising Autoencoders (SDAE) [306], DenseXception block-based CNN [221] (replace DenseNet's convolution layer with Xception layer), CNN-LSTM, denoising autoencoder, LSTM-based multitask learning modality encoder [62], 3D convolutional layers, LSTM	[55,62,87,98,199,221]	Representations, Derived

Table A2. *Cont.*

Features	Tools	Studies	Feature Category
Session-level representations from frame-level features/representations	Average of frame-level representations, Fisher vector (FV) encoding, improved FV coding [265], GMM, Temporal Attentive Pooling (TAP) [75]	[55,75,117,199]	Representations, Derived
Texts extracted from images	python-tesseract [362]	[25,126,128,140]	Textual
Image labels/tags	Deep CNN-based multi-label classifier [113], Contrastive Language Image Pre-training (CLIP) [363], Imagga [364] (CNN-based automatic tagging system)	[113,124,128,129]	Textual
Bag-of-Words (BoVW) features	Multi-scale Dense SIFT features (MSDF) [365]	[124,195]	Textual, Derived
Color distribution-cool, clear, and dominant colors, pixel intensities	Probabilistic Latent Semantic Analysis model [366] (assigns a color to each image pixel), cold color range [367], RGB histogram	[20,140,145,204,220]	Color-related
Brightness, saturation, hue, value, sharpness, contrast, correlation, energy, homogeneity	HSV (Hue, Saturation, color) [368] color model	[20,106,109,113,145,204,220]	Color-related
Statistical descriptors for each HSV distribution: quantiles, mean, variance, skewness, kurtosis	N/A	[145,204]	Color-related, Derived
Pleasure, arousal, and dominance	Compute from brightness and saturation values [276]	[220]	Emotion-related, Derived
Number of pixels, width, height, if image is modified (indicated via exif file)	N/A	[204]	Image metadata

Table A3. Textual features.

Features	Tools	Studies	Feature Category
Count of words: general, condition-specific (depressed, suicidal, eating disorder-related) keywords, emojis	N/A	[20,104,109,123,126,127,130,133,134,137,145,146,187,188,218,219]	Linguistic
Words referring to social processes (e.g., reference to family, friends, social affiliation), and psychological states (e.g., negative/positive emotions)	Linguistic Inquiry and Word Count (LIWC) [278], LIWC 2007 Spanish dictionary [369], Chinese Suicide Dictionary [370], Chinese LIWC [371], TextMind [372], Suite of Automatic Linguistic Analysis Tools (SALAT) [279]—Simple Natural Language Processing (SiNLP) [373]	[20,79,109,118,121,128,186,194,196–198,204,211,219,374]	Linguistic
Part-of-speech (POS) tags: adjectives, nouns, pronouns	Jieba [375], Natural Language Toolkit (NLTK) [280], TextBlob [376], spaCy, Penn Treebank [377], Empath [378]	[61,100,104,123,126,135,184,185,189,195,218,219]	Linguistic

Table A3. *Cont.*

Features	Tools	Studies	Feature Category
Word count-related representations: Term Frequency–Inverse Document Frequency (TF-IDF), Bag of Words (BoW), n-grams, Term Frequency–Category Ratio (TF-CR) [379]	Word2Vec embeddings, language models	[115,116,118,124,128,130, 140,143,144,148,185,186,188, 198,217,374]	Linguistic, Representations
Readability metrics: Automated Readability Index (ARI), Simple Measure of Gobbledygook (SMOG), Coleman–Liau Index (CLI), Flesch reading ease, Gunning fog index, syllable count scores	Textstat [380]	[218,220]	Linguistic
Lexicon-based representations [381]	Depression domain lexicon [382], Chinese suicide dictionary [370]	[120,135,189]	Representations
Sentiment scores, valence, arousal, and dominance (VAD) ratings	NLTK [280], IBM Watson Tone Analyzer, Azure Text Analytics, Google NLP, NRC emotion lexicon [383], senti-py [384], Stanford NLP toolkit [281], Sentiment Analysis and Cognition Engine (SEANCE) [282], text SA API of Baidu Intelligent Cloud Platform [123], Valence Aware Dictionary and Sentiment Reasoner (VADER) [385], Chinese emotion lexicons DUTIR [386], Affective Norms for English Words ratings (ANEW) [283], EmoLex [**?**], SenticNet [388], Lasswell [389], AFINN SA tool [390], LabMT [391], text2emotion [392], BERT [266]	[20,54,61,86,110,115,118,119, 121,123,126–128,130,132,133,137,143–146,148,184–186,188,194,196–198,218,219,374]	Sentiment-related
Happiness scores of emojis	Emoji sentiment scale [393]	[110]	Sentiment-related
Emotion transitions from love to joy, from love to anxiety/sorrow (inspired by [394])	Chinese emotion lexicons DUTIR [386]	[187]	Sentiment-related
Word representations	Global vectors for word representation (GloVe) [395], Word2Vec [396], FastText [397], Embeddings from Language Models (ELMo) [398], BERT [266], ALBERT [297], XLNet [285], bidirectional gated recurrent unit (BiGRU) [341], itwiki (Italian Wikipedia2Vec model), Spanish model [399], EmoBERTa [298] (incorporate linguistic and emotional information), MiniLM [400] (supports multiple languages), GPT [401], TextCNN [402], Bi-LSTM [294]	[49,60,65,67,69,72,73,77,78, 81,82,87,88,90,95–98,100,106,111–113,116,122, 125,128,129,131,135,136,138, 142,145,147,148,185–187,201,214,218,308]	Semantic-related, Representations
Sentence representations	Paragraph Vector (PV) [284], Universal Sentence Encoder [403], Sentence-BERT [404]	[52,59,70,71,89,102,103,174, 199]	Semantic-related, Representations
Topic modeling, topic-level features	Scikit-learn's Latent Dirichlet Allocation module [405], Biterm Topic Model [406]	[20,43,114,118,119,126,130, 134,136,137,146,185,188,194, 217,219]	Semantic-related
Description categories	IBM Watson's Natural Language Understanding tool (https://cloud.ibm.com/apidocs/natural-language-understanding#text-analytics-features (accessed on 10 December 2023))	[132]	Semantic-related
High-level representations from low-level features/representations (e.g., sentence-level from word-level, to capture sequential and/or significant information)	BiLSTM with an attention layer, stacked CNN and BiGRU with attention, summarization [119] using K-means clustering and BART [407], combination of LSTM with attention mechanism and CNN, BiGRU with attention	[73,95,97,119,136,145,159, 201]	Representations, Derived

Table A3. *Cont.*

Features	Tools	Studies	Feature Category
User-level representations from post-level representations	CNN-based triplet network [408] from existing Siamese network [409] (consider cosine similarities between post-level representations between each individual and others in the same and different target groups), LSTM with attention mechanism	[128,138]	Representations, Derived
Session-level representations from segment-level representations	Fisher vector encoding	[199]	Representations, Derived
Subject-level average, median, standard deviation of sentiment scores, representations, POS counts	N/A	[110,185,186]	Derived
Subject-level representations in conversation	Graph attention network—vertex as question/answer pair incorporating LeakyReLU on neighbors with respective attention coefficients, edge between adjacent questions	[97]	Representations, Derived

Table A4. Social media features.

Features	Tools	Studies	Feature Category
Posts distribution (original posts, posts with images, posts of specific emotions/sentiments)-frequency, time	N/A	[109,112,122,123,126,130, 134,137,142,145,188,218,219]	Post metadata
Username, followers, followings, status/bio description, profile header and background images, location, time zone	N/A	[109,115,118,122,123,126, 130,134,137,142,145,171,188, 218,219]	User metadata
Likes, comments, hashtags, mentions, retweets (Twitter), favourites (Twitter)	N/A	[115,126,135,137,142,171, 185,189]	Social interactions, Post metadata
Stressful periods with stress level and category (study, work, family, interpersonal relation, romantic relation, or self-cognition)	Algorithm [410] applied on users' posting behaviors	[187]	Post metadata, Derived
Aggregate posting time by 4 seasons, 7 days of the week, 4 epochs of the day (morning, afternoon, evening, midnight), or specific times (daytime, sleep time, weekdays, weekends)	N/A	[125,130,135,186,188,189, 219]	Post metadata, Derived
Encoding of numerical features	Categorize into quartiles (low, below average, average, high)	[115]	Representations, Derived
Social interaction graph-node: user-level representations concatenated from post-level representations, edge: actions of following, mentioning, replying to comments, quoting	node2vec [411], Ego-network [412]	[139,185]	Social interactions
Personalized graph-user-level node: user-level representations made up of property nodes, property node (individual), personal information, personality, mental health experience, post behavior, emotion expression and social interactions, user–user edge: mutual following-follower relationship, user-property edge: user's characteristics	Attention mechanism to weigh property by contribution to individual's mental health condition (user-property edge) and emotional influence (user–user edge)	[187]	Social interactions
Retweet network node: user-level representations, directed edge: tweets of a user is retweeted by the directed user	Clustering-based neighborhood recognition-form communities with densely connected nodes, expand communities using similarity with adjacent nodes	[141]	Representations

Table A5. Smartphone sensor features.

Features	Tools	Studies	Feature Category
Phone calls and text messages: frequency, duration, entropy	N/A	[104,105,149,155,156,159, 161,162,166,169–171,175,190, 205,206,209,222,223]	Calls and messages
Phone unlocks: frequency, duration	Manual computation, RAPIDS [413]-a tool for data pre-processing and biomarker computation	[99,149,150,155,156,158,158, 160–162,166,167,171,176,190, 205,206,208,212,374]	Phone interactions
Phone charge duration	N/A	[163]	Phone interactions
Running applications: type, frequency, duration of usage	N/A	[99,149,150,155,156,158,160–162,166,169–171,190,205,206,208,212,374]	Phone interactions
Activity states (e.g., walking, stationary, exercising, running, unknown): frequency, duration	Android activity recognition API, activity recognition model (LSTM-RNN [414], SVM), Google Activity Recognition Transition API (using gyroscope and accelerometer)	[150,152,154,160,163,169, 170,176,177,190,205,206]	Physical mobility
Distance traveled, displacement from home, location variance and entropy, time spent at specific places, transitions	Manual computation, RAPIDS [413]	[99,150,151,153–155,158,160–162,165,166,175–177,200,205,206,208,209]	Physical mobility
Location cluster features: number of clusters, largest cluster as primary location, most and least visited clusters	DBSCAN clustering [415], Adaptive K-means clustering [416]	[150,151,153,154,160,165, 176,177,205,208]	Physical mobility
Speed	Compute from GPS and/or accelerometer	[153,165,166,209]	Physical mobility
Intensity of action	Compute rotational momentum from GPS and gyroscope	[162]	Physical mobility
GPS sensor, calls and phone screen unlock features	RAPIDS [413]-a tool for data pre-processing and biomarker computation	[158,164]	Physical mobility, Calls and messages, Phone interactions
WiFi association events (when a smartphone is associated or dissociated with a nearby access point at a location's WiFi network)	N/A	[153]	Connectivity
Occurrences of unique Bluetooth addresses, most/least frequently detected devices	N/A	[99,151,155,156,175]	Connectivity
Surrounding sound: amplitude, conversations, human/non-human voices	N/A	[150,163,166,205–209]	Ambient environment
Surrounding illuminance: amplitude, mean, variance, standard deviation	N/A	[99,163,190,205,208,209]	Ambient environment
Silent and noise episodes: count, sum, minimum decibels	Detect via intermittent samples until noise state changes	[166]	Ambient environment
Sleep duration, wake and sleep onset	Infer from ambient light, audio amplitude, activity state, and screen on/off	[150,160,161,167,169,170, 175,176,206]	Derived, Physical mobility
Keystroke features: count, transitions, time between two consecutive keystrokes	N/A	[166,202]	Phone interactions
Time between two successive touch interactions (tap, long tap, touch)	N/A	[166]	Phone interactions
Day-level features	Statistical functions (mean, median, mode, standard deviation, interquartile range) at the day-level or day of the week (weekdays, weekends)	[151,152,154,156,159,163, 164,170,176,206]	Derived

Table A5. *Cont.*

Features	Tools	Studies	Feature Category
Epoch-level features	Statistical functions at partitions of a day-morning, afternoon, evening, night	[149,151,152,156,159,163,166,176,206]	Derived
Hour-level features	Statistical functions at each hour of the day	[208,209]	Derived
Week-level features	Statistical functions at the week-level, distance from weekly mean	[162,164]	Derived
Rhythm-related features: ultradian, circadian, and infradian rhythms, regularity index [417], periodicity based on time windows	Manual computation, Cosinor [418]-a rhythmic regression function	[151–153,155,157,158,176,207]	Derived
Degrees of complexity and irregularity	Shannon entropy of sensor features	[166]	Derived
Statistical, temporal and spectral time series features	Time Series Feature Extraction Library (TSFEL) [419]	[104,105]	Derived
High-level cluster-based features: cluster labels, likelihood scores, distance scores, transitions	Gaussian mixture model (GMM) [420], partition around methods (PAM) clustering model [421]	[208,209]	Derived
Network of social interactions and personal characteristics: node type corresponds to a modality/category (e.g., individual, personality traits, social status, physical health, well-being, mental health status)	Heterogeneous Information Network (HIN) [422]	[173]	Representations
Representations capturing important patterns across timestamps	Transformer encoder [295]	[179]	Representations

Table A6. Wearable sensor features.

Features	Tools	Studies	Feature Category
Duration and onset of sleep status (asleep, restless, awake, unknown), sleep efficiency, sleep debt	API of wristband	[149,151,155,156,164,171,180–182,191,374]	Physical mobility
Number of steps, active and sedentary bouts, floor climb	API of wristband	[150,151,155,156,164,171,179–182,191,374]	Physical mobility
Heart rate (HR), galvanic skin response (GSR), skin temperature (ST), electrodermal activity (EDA)	API of Wristband	[149,150,164,169,170,172,178,179,182,191]	Physiological
Outliers of systolic and diastolic periods: centering tendency, spreading degree, distribution shape and symmetry degree values from blood volume pressure	N/A	[178]	Physiological, Derived
Motion features from accelerometer data: acceleration, motion	N/A	[149]	Physical mobility
Heart rate variability (HRV), rapid eye movement, wake after sleep onset, metabolic equivalent for task (MET) for physical activity	API of Oura ring	[158]	Physiological, Physical mobility
High-level features from HR, GSR, and ST signals	CNN-LSTM	[215]	Representations
Basal metabolic rate (BMR) calories	API of wristband	[179,180]	Physiological

Table A7. Demographic and personality features.

Features	Tools	Studies	Feature Category
Gender, age, location	Sina microblog user account	[187]	Demographic
Gender, age, relationships, education levels	bBridge [423], big data platform for social multimedia analytics	[20]	Demographic
Age, gender	Age and gender lexica [424], M3-inference model [425] performs multimodal analysis on profile images, usernames, and descriptions on social media profiles	[121,143,144]	Demographic
Big 5 personality scores	IBM's Personality Insights [426], BERT-MLP model [427] on textual content	[57,121,130,143,144,188]	Personality
Proportion of perfection and ruminant thinking-related words in textual content (inspired by [287])	Perfection and ruminant-thinking-related lexicons	[187]	Personality
Interpersonal sensitivity: amount of stressful periods associated with interpersonal relations	Algorithm [410] applied on users' posting behaviors	[187]	Personality

Appendix B. Existing Modality Fusion Techniques

Table A8. Modality fusion techniques.

Category	Method	Tools	Studies
Feature level	Concatenate into a single representation	N/A	[67,84,85,89,96,97,105,132,142,143,145,146,166,170,179,197,199–201,217]
Score/Decision level	Sum-rule, product-rule, max-rule, AND and OR operations, or majority voting on modality-level scores	N/A	[48,51,56,77,87,98,126,173,193,198,201]
	Weighted average or sum of modality-level scores	N/A	[51,68,147,198,200]
	Average confidence scores from lower-level prediction	N/A	[121]
	Combine predictions of individual modalities as inputs to secondary ML models	SVM, decision tree, random forest, novel ML models	[48,52,56,64,71,72,74,103,122,155,193]
	Hierarchical score/decision-level fusion	Weighted voting fusion network [428]	[122,195]
	Summation of question-level scores from rules enforced on modality-specific predictions	N/A	[88]
Model level	Map multiple features into a single vector	LSTM-based encoder–decoder network, LSTM-based neural network, BiLSTM, LSTM, fully connected layer, tensor fusion network	[46,59,75,80,86,95,187]
	Concatenate feature representations as a single input to learn high-level representations	Dense and fully connected layers with attention mechanisms, CNN, multi-head attention network, transformer [295], novel time-aware LSTM	[70,73,77,89,91,92,94,125,189,214]

Table A8. *Cont.*

Category	Method	Tools	Studies
Model level	Learn shared representations from weighted modality-specific representations	Gated Multimodal Unit (GMU) [429], parallel attention model, attention layer, sparse MLP (mix vertical and horizontal information via weight sharing and sparse connection), multimodal encoder–decoder, multimodal factorized bilinear pooling (combines compact output features of multi-modal low-rank bilinear [430] and robustness of multi-modal compact bilinear [431]), multi-head intermodal attention fusion, transformer [295], feed-forward network, low-rank multimodal fusion network [432]	[62,65,67,76,93,100,102,106,113, 117,131,135,136,142–144,174,218,433]
	Learn joint sparse representations	Dictionary learning	[20]
	Learn and fuse outputs from different modality-specific parts at fixed time steps	Cell-coupled LSTM with L-skip fusion mechanism	[101]
	Learn cross-modality representations that incorporate interactions between modalities	LXMERT [434], transformer encoder with cross-attention layers (representations of a modality as query and the other as key/value, and vice versa), memory fusion network [435]	[82,92,129]
	Horizontal and vertical kernels to capture patterns across different levels	CASER [309]	[170]

Appendix C. Existing Machine Learning Models

Table A9. Machine learning models.

Category	Machine Learning Models	Application Method	Studies
Supervised learning	Linear regression, logistic regression, least absolute shrinkage and selection operator (Lasso) regularized linear regression [436], ElasticNet regression [437], stochastic gradient descent (SGD) regression, Gaussian staircase model, partial least square (PLS) [438] regression (useful for collinear features), generalized linear models	Learn relationship between features to predict continuous values (scores of assessment scales) or probabilities (correspond to output classes)	[20,43,49,53,55,68,70,99,104,105,126,130,134, 140,150,154,163,164,167,175,179,182,188,200, 211–213,219,222,223]
	SVM	Find a hyperplane that best fits features (regression) or divides features into classes (classification), secondary model in score-level fusion	[47,50,79,99,104,105,115,121,130,134,140,148, 162,163,169,178,179,188,198,210,219,223]
	One class SVM [439]	Anomaly detection by treating outliers as points on the other side of hyperplane	[165]
	Three-step hierarchical logistic regression	Incremental inclusion of three feature groups in conventional logistic regression	[181]
	Discriminant functions-Naive Bayes, quadratic discriminant analysis (QDA), linear discriminant analysis (LDA), Gaussian naive Bayes	Determine class based on Bayesian probabilities, detect state changes	[12,99,104,140,148,152,163,222]
	Decision tree	Construct a tree that splits into leaf nodes based on feature	[99,134,140,148,164,178]
	Mixed-effect classification and regression trees-generalized linear mixed-effects model (GLMM) trees [440]	Capture interactions and nonlinearity among features while accounting for longitudinal structure	[191]

Table A9. *Cont.*

Category	Machine Learning Models	Application Method	Studies
Neural network	Fully connected (FC) layers, multilayer perceptron (MLP), CNN, LSTM, BiLSTM, GRU, temporal convolutional network (TCN) [441] (with dilation for long sequences)-with activation function like Sigmoid, Softmax, ReLU, LeakyReLU, and GeLU	Predict scores of assessment scales (regression) or probability distribution over classes (classification)	[60,78,80,84–88,90–94,96,98,105,111,113,117,131,133,135,136,142–144,146,162,163,167,168,170,172,174,178,179,190,197,199,201,218,219,221,223,308]
	DCNN-DNN (combination of deep CNN and DNN), GCNN-LSTM (combination of gated convolutional neural network, which replaces a convolution block in CNN with a gated convolution block, and LSTM)	The latter neural network makes predictions based on high-level global features learned by the prior	[52,308]
	Cross-domain DNN with feature adaptive transformation and combination strategy (DNN-FATC)	Enhance detection in the target domain by transferring information from a heterogenous source domain	[109]
	Attention-based TCN	Classify features using relational classification attention [442]	[72]
	One-hot transformer (lower complexity than original sine and cosine functions)	Apply one-hot encoding on features for classification	[72]
	Transformer [295]	Apply self-attention across post-level representations, attention masking masks missing information	[129]
	Transformer-based sequence classification models-BERT, RoBERTa [296], XLNet [285], Informer [443] (for long sequences)	Perform classification using custom pre-trained tokenizers augmented with special tokens for tokenization	[121,179]
	Hierarchical attention network (HAN) [444]	Predict on user-level representations derived from stacked attention-based post-level representations, each made up of attention-based word-level representations	[128]
	LSTM-based encoder and decoder	Learn factorized joint distributions to generate modality-specific generative factors and multimodal discriminative factors to reconstruct unimodal inputs and predict labels respectively	[82]
	GRU-RNN as baseline model with FC layers as personalized model	Train baseline model using data from all samples and fine-tune personalized model on individual samples	[161]
	CNN-based triplet network [408]	Incorporate representations of homogeneous users	[138]
	Stacked graph convolutional network	Perform classification on heterogeneous graphs by learning embeddings, sorting graph nodes, and performing graph comparisons	[139]
	GRU-D (introduce decay rates in conventional GRU to control decay mechanism)	Learn feature-specific hidden decay rates from inputs	[171]
Ensemble learning	Random forest (RF) [300], eXtreme Gradient Boosting (XGBoost), AdaBoost [301], Gradient Boosted Regression Tree [302] (GDBT) (less sensitive to outliers and more robust to overfitting)	Predict based on numerical input features	[51,99,104,105,114,126,130,134,140,148,151,155,157,160,163,164,167,169,178,179,182,183,188,203,204,206,212,219,222,223]
	RF	Secondary model that predicts from regression scores and binary outputs of individual modality predictions	[71,81]
	Balanced RF [445] (RF on imbalanced data)	Aggregate predictions of ensemble on balanced down-sampled data	[209]
	XGBoost-based subject-specific hierarchical recall network	Deduce subject-level labels based on whether the output probability of XGBoost at a specific layer exceeds a predetermined threshold	[194]
	Stacked ensemble learning architecture	Obtain the first level of predictions from KNN, naive Bayes, Lasso regression, ridge regression, and SVM, then use them as features of a second-layer logistic regression	[123]
	Feature-stacking (a meta-learning approach) [303]	Use logistic regression as an L1 learner to combine predictions of weak L0 learners on different feature sets	[185]

Table A9. *Cont.*

Category	Machine Learning Models	Application Method	Studies
Ensemble learning	Greedy Ensembles of Weighted Extreme Learning Machines (GEWELMs), WELM [446] (weighted mapping for unbalanced class), Kernel ELM	ELM [447] as a building block that maps inputs to class-based outputs via least square regression	[63,127,192]
	Stacked ensemble classifier	Use MLP as meta learner to integrate outputs of CNN base learners	[126]
	Cost-sensitive boosting pruning trees-AdaBoost with pruned decision trees	Weighted pruning prunes redundant leaves to increase generalization and robustness	[137]
	Weighted voting model	Weight predictions of baseline ML models (DT, Naive Bayes, KNN, SVM, generalized linear models, GDBT) based on class probabilities and deduce final outcome from the highest weighted class	[140]
	Ensemble of SVM, DT, and naive Bayes	N/A	[89]
	Combination of personalized LSTM-based and RF models	Train personalized LSTM on hourly time series data (of another sample most similar to the sample of concern based on demographic characteristics and baseline MH states), and RF on statistical and cluster-based features	[208]
Multi-task learning	CNN	Train jointly to produce two output branches, regression score and probability distribution for classification	[61,62]
	LSTM-RNN, attention-based LSTM subnetwork, MLP with shared and task-specific layers	Train for depression prediction with emotion recognition as the secondary task	[46,106,132]
	LSTM with Swish [448] activation function (speeds up training with the advantages of linear and ReLU activation), GRU with FC layers, DNN with multi-task loss function	Perform both regression and classification simultaneously	[74,102,118,141,193]
	Multi-task FC layers	Train jointly to predict severity level and discrete probability distribution	[97]
	Multi-output support least-squares vector regression machines (m-SVR) [304]	Map multivariate inputs to a multivariate output space to predict several tasks	[207]
	2-layer MLP with shared and task-specific dense layers with dynamic weight tuning technique	Train to perform individual predictions for positive and control groups	[180]
	Bi-LSTM-based DNNs to provide auxiliary outputs into DNN for main output	Auxiliary outputs correspond to additional predictions to incorporate additional information	[176]
	DNN (FC layers with Softmax activation) for auxiliary and main outputs	Train DNNs individually on different feature combinations as individual tasks to obtain auxiliary losses for joint optimization function of main output	[145]
	Multi-task neural network with shared LSTM layer and two task-specific LSTM layers	Train to predict male and female samples individually	[70]
Others	Semi-supervised learning-ladder network classifier [305] of stacked noisy encoder and denoising autoencoder [306]	Reconstruct input using outputs of noisy encoder in the current layer and decoder from the previous layer, combine with MLP (inspired by [449])	[196]
	DMF [450], RESCAL [451], DEDICOM [452], HERec [453]	Perform recommender system [307] approach on features modeled using HIN	[173]
	Graphlets [454], colored graphlets [455], DeepWalk [456], Metapath2vec++ [457]	Perform node classification on features modeled using HIN	[173]
	Combination of DBSCAN and K-Means	Density-based clustering	[78]
	Clustering-based-KNN	Deduce predicted class through voting of K-nearest data	[140,163,166,178,212,223]
	Linear superimpose of modality-specific features	Learn fitting parameters (between 0 and 1) that adjust the proportions of modality-specific features in the final outcome	[83]
	Two-staged prediction with outlier detection	Baseline ML model (LR, SVM, KNN, DT, GBDT, AdaBoost, RF, Gaussian naive Bayes, LDA, QDA, DNN, CNN) performs day-level predictions, t-test detects outliers in first stage outputs	[163]

Table A9. *Cont.*

Category	Machine Learning Models	Application Method	Studies
Others	Label association mechanism	Apply to one-hot vectors of predictions from modality-specific DNNs	[189]
	Isolation Forest (ISOFOR) [458], Local Outlier Factor (LOF) [459], Connectivity-Based Outlier Factor (COF) [460]	Unsupervised anomaly detection	[166]
	Similarity and threshold relative to the model of normality (MoN) (from the average of deep representations of training instances in respective target groups)	Deduce predicted class based on higher similarity with corresponding MoN	[85]
	Federated learning based on DNN	Train global model on all data and fine-tune the last layer locally	[168]

References

1. Institute of Health Metrics and Evaluation. *Global Health Data Exchange (GHDx)*; Institute of Health Metrics and Evaluation: Seattle, WA, USA, 2019.
2. World Health Organization. *Mental Health and COVID-19: Early Evidence of the Pandemic's Impact: Scientific Brief, 2 March 2022*; Technical Report; World Health Organization: Geneva, Switzerland, 2022.
3. Australian Bureau of Statistics (2020–2022). National Study of Mental Health and Wellbeing. 2022. Available online: https://www.abs.gov.au/statistics/health/mental-health/national-study-mental-health-and-wellbeing/latest-release (accessed on 10 December 2023).
4. National Institute of Mental Health. Statistics of Mental Illness. 2021. Available online: https://www.nimh.nih.gov/health/statistics/mental-illness (accessed on 10 December 2023).
5. Bloom, D.; Cafiero, E.; Jané-Llopis, E.; Abrahams-Gessel, S.; Bloom, L.; Fathima, S.; Feigl, A.; Gaziano, T.; Hamandi, A.; Mowafi, M.; et al. *The Global Economic Burden of Noncommunicable Diseases*; Technical Report; Harvard School of Public Health: Boston, MA, USA, 2011.
6. World Health Organization. Mental Health and Substance Use. In *Comprehensive Mental Health Action Plan 2013–2030*; World Health Organization: Geneva, Switzerland, 2021.
7. Borg, M. The Nature of Recovery as Lived in Everyday Life: Perspectives of Individuals Recovering from Severe Mental Health Problems. Ph.D. Thesis, Norwegian University of Science and Technology, Trondheim, Norway, 2007.
8. Barge-Schaapveld, D.Q.; Nicolson, N.A.; Berkhof, J.; Devries, M.W. Quality of life in depression: Daily life determinants and variability. *Psychiatry Res.* **1999**, *88*, 173–189. [CrossRef] [PubMed]
9. Rapee, R.M.; Heimberg, R.G. A cognitive-behavioral model of anxiety in social phobia. *Behav. Res. Ther.* **1997**, *35*, 741–756. [CrossRef]
10. Stewart-Brown, S. Emotional wellbeing and its relation to health. *BMJ* **1998**, *317*, 1608–1609.
11. Goldman, L.S.; Nielsen, N.H.; Champion, H.C. The Council on American Medical Association Council on Scientific Affairs. Awareness, Diagnosis, and Treatment of Depression. *J. Gen. Intern. Med.* **1999**, *14*, 569–580. [CrossRef]
12. Grünerbl, A.; Muaremi, A.; Osmani, V.; Bahle, G.; Öhler, S.; Tröster, G.; Mayora, O.; Haring, C.; Lukowicz, P. Smartphone-Based Recognition of States and State Changes in Bipolar Disorder Patients. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 140–148. [CrossRef]
13. Kakuma, R.; Minas, H.; Ginneken, N.; Dal Poz, M.; Desiraju, K.; Morris, J.; Saxena, S.; Scheffler, R. Human resources for mental health care: Current situation and strategies for action. *Lancet* **2011**, *378*, 1654–1663. [CrossRef]
14. Le Glaz, A.; Haralambous, Y.; Kim-Dufor, D.H.; Lenca, P.; Billot, R.; Ryan, T.C.; Marsh, J.; DeVylder, J.; Walter, M.; Berrouiguet, S.; et al. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. *J. Med. Internet Res.* **2021**, *23*, e15708. [CrossRef]
15. Rahman, R.A.; Omar, K.; Mohd Noah, S.A.; Danuri, M.S.N.M.; Al-Garadi, M.A. Application of Machine Learning Methods in Mental Health Detection: A Systematic Review. *IEEE Access* **2020**, *8*, 183952–183964. [CrossRef]
16. Graham, S.; Depp, C.; Lee, E.E.; Nebeker, C.; Tu, X.; Kim, H.C.; Jeste, D.V. Artificial Intelligence for Mental Health and Mental Illnesses: An Overview. *Curr. Psychiatry Rep.* **2019**, *21*, 116. [CrossRef]
17. Thieme, A.; Belgrave, D.; Doherty, G. Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Trans. Comput. Hum. Interact.* **2020**, *27*, 1–53. [CrossRef]
18. Javaid, M.; Haleem, A.; Pratap Singh, R.; Suman, R.; Rab, S. Significance of machine learning in healthcare: Features, pillars and applications. *Int. J. Intell. Netw.* **2022**, *3*, 58–73. [CrossRef]
19. Riaz Choudhry, F.; Vasudevan Mani, L.C.M.; Khan, T.M. Beliefs and perception about mental health issues: A meta-synthesis. *Neuropsychiatr. Dis. Treat.* **2016**, *12*, 2807–2818. [CrossRef] [PubMed]
20. Shen, G.; Jia, J.; Nie, L.; Feng, F.; Zhang, C.; Hu, T.; Chua, T.S.; Zhu, W. Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 3838–3844.
21. Manickam, P.; Mariappan, S.A.; Murugesan, S.M.; Hansda, S.; Kaushik, A.; Shinde, R.; Thipperudraswamy, S.P. Artificial Intelligence (AI) and Internet of Medical Things (IoMT) Assisted Biomedical Systems for Intelligent Healthcare. *Biosensors* **2022**, *12*, 562. [CrossRef]

22. Skaik, R.; Inkpen, D. Using Social Media for Mental Health Surveillance: A Review. *ACM Comput. Surv.* **2020**, *53*, 1–31. [CrossRef]
23. Chen, X.; Genc, Y. A Systematic Review of Artificial Intelligence and Mental Health in the Context of Social Media. In Proceedings of the Artificial Intelligence in HCI, Virtual, 26 June–1 July 2022; pp. 353–368.
24. Deshmukh, V.M.; Rajalakshmi, B.; Dash, S.; Kulkarni, P.; Gupta, S.K. Analysis and Characterization of Mental Health Conditions based on User Content on Social Media. In Proceedings of the 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 28–29 January 2022; pp. 1–5. [CrossRef]
25. Yazdavar, A.H.; Mahdavinejad, M.S.; Bajaj, G.; Romine, W.; Sheth, A.; Monadjemi, A.H.; Thirunarayan, K.; Meddar, J.M.; Myers, A.; Pathak, J.; et al. Multimodal mental health analysis in social media. *PLoS ONE* **2020**, *15*, e0226248. [CrossRef]
26. Garcia Ceja, E.; Riegler, M.; Nordgreen, T.; Jakobsen, P.; Oedegaard, K.; Torresen, J. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive Mob. Comput.* **2018**, *51*, 1–26. [CrossRef]
27. Hickey, B.A.; Chalmers, T.; Newton, P.; Lin, C.T.; Sibbritt, D.; McLachlan, C.S.; Clifton-Bligh, R.; Morley, J.; Lal, S. Smart Devices and Wearable Technologies to Detect and Monitor Mental Health Conditions and Stress: A Systematic Review. *Sensors* **2021**, *21*, 3461. [CrossRef]
28. Woodward, K.; Kanjo, E.; Brown, D.J.; McGinnity, T.M.; Inkster, B.; Macintyre, D.J.; Tsanas, A. Beyond Mobile Apps: A Survey of Technologies for Mental Well-Being. *IEEE Trans. Affect. Comput.* **2022**, *13*, 1216–1235. [CrossRef]
29. Craik, K.H. The lived day of an individual: A person-environment perspective. *Pers. Environ. Psychol. New Dir. Perspect.* **2000**, *2*, 233–266.
30. Harari, G.M.; Müller, S.R.; Aung, M.S.; Rentfrow, P.J. Smartphone sensing methods for studying behavior in everyday life. *Curr. Opin. Behav. Sci.* **2017**, *18*, 83–90. [CrossRef]
31. Stucki, R.A.; Urwyler, P.; Rampa, L.; Müri, R.; Mosimann, U.P.; Nef, T. A Web-Based Non-Intrusive Ambient System to Measure and Classify Activities of Daily Living. *J. Med. Internet Res.* **2014**, *16*, e175. [CrossRef]
32. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, 105906. [CrossRef]
33. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ* **2009**, *339*, W-65–W-94. [CrossRef]
34. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Group, T.P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **2009**, *6*, 336–341. [CrossRef] [PubMed]
35. Kitchenham, B.; Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Technical Report; University of Durham: Durham, UK, 2007.
36. Zhang, T.; Schoene, A.; Ji, S.; Ananiadou, S. Natural language processing applied to mental illness detection: A narrative review. *npj Digit. Med.* **2022**, *5*, 46. [CrossRef]
37. Valstar, M.; Schuller, B.; Smith, K.; Eyben, F.; Jiang, B.; Bilakhia, S.; Schnieder, S.; Cowie, R.; Pantic, M. AVEC 2013: The Continuous Audio/Visual Emotion and Depression Recognition Challenge. In Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge (AVEC ’13), Barcelona, Spain, 21 October 2013; pp. 3–10. [CrossRef]
38. Valstar, M.; Schuller, B.; Smith, K.; Almaev, T.; Eyben, F.; Krajewski, J.; Cowie, R.; Pantic, M. AVEC 2014: 3D Dimensional Affect and Depression Recognition Challenge. In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC ’14), Orlando, FL, USA, 7 November 2014; pp. 3–10. [CrossRef]
39. Sawyer, S.M.; Azzopardi, P.S.; Wickremarathne, D.; Patton, G.C. The age of adolescence. *Lancet Child Adolesc. Health* **2018**, *2*, 223–228. [CrossRef]
40. Semrud-Clikeman, M.; Goldenring Fine, J. Pediatric versus adult psychopathology: Differences in neurological and clinical presentations. In *The Neuropsychology of Psychopathology*; Contemporary Neuropsychology; Springer: New York, NY, USA, 2013; pp. 11–27.
41. Cobham, V.E.; McDermott, B.; Haslam, D.; Sanders, M.R. The Role of Parents, Parenting and the Family Environment in Children’s Post-Disaster Mental Health. *Curr. Psychiatry Rep.* **2016**, *18*, 53. [CrossRef]
42. Tuma, J.M. Mental health services for children: The state of the art. *Am. Psychol.* **1989**, *44*, 188–199. [CrossRef]
43. Gong, Y.; Poellabauer, C. Topic Modeling Based Multi-Modal Depression Detection. In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (AVEC ’17), Mountain View, CA, USA, 23 October 2017; pp. 69–76. [CrossRef]
44. Van Praag, H. Can stress cause depression? *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2004**, *28*, 891–907. [CrossRef]
45. Power, M.J.; Tarsia, M. Basic and complex emotions in depression and anxiety. *Clin. Psychol. Psychother.* **2007**, *14*, 19–31. [CrossRef]
46. Chao, L.; Tao, J.; Yang, M.; Li, Y. Multi task sequence learning for depression scale prediction from video. In Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi’an, China, 21–24 September 2015; pp. 526–531. [CrossRef]
47. Yang, L.; Jiang, D.; He, L.; Pei, E.; Oveneke, M.C.; Sahli, H. Decision Tree Based Depression Classification from Audio Video and Language Information. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge (AVEC ’16), Amsterdam, The Netherlands, 16 October 2016; pp. 89–96. [CrossRef]

48. Pampouchidou, A.; Simantiraki, O.; Fazlollahi, A.; Pediaditis, M.; Manousos, D.; Roniotis, A.; Giannakakis, G.; Meriaudeau, F.; Simos, P.; Marias, K.; et al. Depression Assessment by Fusing High and Low Level Features from Audio, Video, and Text. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge (AVEC '16), Amsterdam, The Netherlands, 16 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 27–34. [CrossRef]
49. Williamson, J.R.; Godoy, E.; Cha, M.; Schwarzentruber, A.; Khorrami, P.; Gwon, Y.; Kung, H.T.; Dagli, C.; Quatieri, T.F. Detecting Depression Using Vocal, Facial and Semantic Communication Cues. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge (AVEC '16), Amsterdam, The Netherlands, 16 October 2016; pp. 11–18. [CrossRef]
50. Smailis, C.; Sarafianos, N.; Giannakopoulos, T.; Perantonis, S. Fusing Active Orientation Models and Mid-Term Audio Features for Automatic Depression Estimation. In Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '16), Corfu Island, Greece, 29 June–1 July 2016; Association for Computing Machinery: New York, NY, USA, 2016. [CrossRef]
51. Nasir, M.; Jati, A.; Shivakumar, P.G.; Nallan Chakravarthula, S.; Georgiou, P. Multimodal and Multiresolution Depression Detection from Speech and Facial Landmark Features. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge (AVEC '16), Amsterdam, The Netherlands, 16 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 43–50. [CrossRef]
52. Yang, L.; Jiang, D.; Xia, X.; Pei, E.; Oveneke, M.C.; Sahli, H. Multimodal Measurement of Depression Using Deep Learning Models. In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (AVEC '17), Mountain View, CA, USA, 23–27 October 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 53–59. [CrossRef]
53. Jan, A.; Meng, H.; Gaus, Y.F.B.A.; Zhang, F. Artificial Intelligent System for Automatic Depression Level Analysis Through Visual and Vocal Expressions. *IEEE Trans. Cogn. Dev. Syst.* **2018**, *10*, 668–680. [CrossRef]
54. Samareh, A.; Jin, Y.; Wang, Z.; Chang, X.; Huang, S. Detect depression from communication: How computer vision, signal processing, and sentiment analysis join forces. *IISE Trans. Healthc. Syst. Eng.* **2018**, *8*, 196–208. [CrossRef]
55. Dibeklioğlu, H.; Hammal, Z.; Cohn, J.F. Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 525–536. [CrossRef]
56. Alghowinem, S.; Goecke, R.; Wagner, M.; Epps, J.; Hyett, M.; Parker, G.; Breakspear, M. Multimodal Depression Detection: Fusion Analysis of Paralinguistic, Head Pose and Eye Gaze Behaviors. *IEEE Trans. Affect. Comput.* **2018**, *9*, 478–490. [CrossRef]
57. Kim, J.Y.; Kim, G.Y.; Yacef, K. Detecting Depression in Dyadic Conversations with Multimodal Narratives and Visualizations. In Proceedings of the AI 2019: Advances in Artificial Intelligence, Adelaide, Australia, 2–5 December 2019; Liu, J., Bailey, J., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 303–314.
58. Victor, E.; Aghajan, Z.M.; Sewart, A.R.; Christian, R. Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation. *Psychol. Assess.* **2019**, *31*, 1019–1027. [CrossRef]
59. Ray, A.; Kumar, S.; Reddy, R.; Mukherjee, P.; Garg, R. Multi-Level Attention Network Using Text, Audio and Video for Depression Prediction. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19), Nice, France, 21–25 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 81–88. [CrossRef]
60. Rodrigues Makiuchi, M.; Warnita, T.; Uto, K.; Shinoda, K. Multimodal Fusion of BERT-CNN and Gated CNN Representations for Depression Detection. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19), Nice, France, 21–25 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 55–63. [CrossRef]
61. Fan, W.; He, Z.; Xing, X.; Cai, B.; Lu, W. Multi-Modality Depression Detection via Multi-Scale Temporal Dilated CNNs. In Proceedings of the Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19), Nice, France, 21–25 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 73–80. [CrossRef]
62. Qureshi, S.A.; Saha, S.; Hasanuzzaman, M.; Dias, G. Multitask Representation Learning for Multimodal Estimation of Depression Level. *IEEE Intell. Syst.* **2019**, *34*, 45–52. [CrossRef]
63. Kaya, H.; Fedotov, D.; Dresvyanskiy, D.; Doyran, M.; Mamontov, D.; Markitantov, M.; Akdag Salah, A.A.; Kavcar, E.; Karpov, A.; Salah, A.A. Predicting Depression and Emotions in the Cross-Roads of Cultures, Para-Linguistics, and Non-Linguistics. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19), Nice, France, 21–25 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 27–35. [CrossRef]
64. Muszynski, M.; Zelazny, J.; Girard, J.M.; Morency, L.P. Depression Severity Assessment for Adolescents at High Risk of Mental Disorders. In Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20), Virtual, 25–29 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 70–78. [CrossRef]
65. Aloshban, N.; Esposito, A.; Vinciarelli, A. Detecting Depression in Less Than 10 Seconds: Impact of Speaking Time on Depression Detection Sensitivity. In Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20), Virtual, 25–29 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 79–87. [CrossRef]
66. Liu, Z.; Wang, D.; Ding, Z.; Chen, Q. A Novel Bimodal Fusion-based Model for Depression Recognition. In Proceedings of the 2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM), Shenzhen, China, 1–2 March 2021; pp. 1–4. [CrossRef]
67. Toto, E.; Tlachac, M.; Rundensteiner, E.A. AudiBERT: A Deep Transfer Learning Multimodal Classification Framework for Depression Screening. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM '21), Queensland, Australia, 1–5 November 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 4145–4154. [CrossRef]

68. Chordia, A.; Kale, M.; Mayee, M.; Yadav, P.; Itkar, S. Automatic Depression Level Analysis Using Audiovisual Modality. In *Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics*; Satapathy, S.C., Bhateja, V., Favorskaya, M.N., Adilakshmi, T., Eds.; Springer: Singapore, 2021; pp. 425–439.
69. Muzammel, M.; Salam, H.; Othmani, A. End-to-end multimodal clinical depression recognition using deep neural networks: A comparative analysis. *Comput. Methods Programs Biomed.* **2021**, *211*, 106433. [CrossRef]
70. Oureshi, S.A.; Dias, G.; Saha, S.; Hasanuzzaman, M. Gender-Aware Estimation of Depression Severity Level in a Multimodal Setting. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8. [CrossRef]
71. Yang, L.; Jiang, D.; Sahli, H. Integrating Deep and Shallow Models for Multi-Modal Depression Analysis—Hybrid Architectures. *IEEE Trans. Affect. Comput.* **2021**, *12*, 239–253. [CrossRef]
72. Ye, J.; Yu, Y.; Wang, Q.; Li, W.; Liang, H.; Zheng, Y.; Fu, G. Multi-modal depression detection based on emotional audio and evaluation text. *J. Affect. Disord.* **2021**, *295*, 904–913. [CrossRef]
73. Shen, Y.; Yang, H.; Lin, L. Automatic Depression Detection: An Emotional Audio-Textual Corpus and A Gru/Bilstm-Based Model. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Virtual, 1–3 May 2022; pp. 6247–6251. [CrossRef]
74. Liu, J.; Huang, Y.; Chai, S.; Sun, H.; Huang, X.; Lin, L.; Chen, Y.W. Computer-Aided Detection of Depressive Severity Using Multimodal Behavioral Data. In *Handbook of Artificial Intelligence in Healthcare: Advances and Applications*; Springer: Cham, Switzerland, 2022; Volume 1, pp. 353–371. [CrossRef]
75. Uddin, M.A.; Joolee, J.B.; Sohn, K.A. Deep Multi-Modal Network Based Automated Depression Severity Estimation. *IEEE Trans. Affect. Comput.* **2022**, *14*, 2153–2167. [CrossRef]
76. Cao, Y.; Hao, Y.; Li, B.; Xue, J. Depression prediction based on BiAttention-GRU. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *13*, 5269–5277. [CrossRef]
77. Mao, K.; Zhang, W.; Wang, D.B.; Li, A.; Jiao, R.; Zhu, Y.; Wu, B.; Zheng, T.; Qian, L.; Lyu, W.; et al. Prediction of Depression Severity Based on the Prosodic and Semantic Features with Bidirectional LSTM and Time Distributed CNN. *IEEE Trans. Affect. Comput.* **2022**, *14*, 2251–2265. [CrossRef]
78. Aloshban, N.; Esposito, A.; Vinciarelli, A. What You Say or How You Say It? Depression Detection Through Joint Modeling of Linguistic and Acoustic Aspects of Speech. *Cogn. Comput.* **2021**, *14*, 1585–1598. [CrossRef]
79. Bilalpur, M.; Hinduja, S.; Cariola, L.A.; Sheeber, L.B.; Alien, N.; Jeni, L.A.; Morency, L.P.; Cohn, J.F. Multimodal Feature Selection for Detecting Mothers' Depression in Dyadic Interactions with their Adolescent Offspring. In Proceedings of the 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), Waikoloa Beach, HI, USA, 5–8 January 2023; pp. 1–8. [CrossRef]
80. Flores, R.; Tlachac, M.; Toto, E.; Rundensteiner, E. AudiFace: Multimodal Deep Learning for Depression Screening. In Proceedings of the 7th Machine Learning for Healthcare Conference (PMLR), Durham, NC, USA, 5–6 August 2022; Proceedings of Machine Learning Research; Lipton, Z., Ranganath, R., Sendak, M., Sjoding, M., Yeung, S., Eds.; 2022; Volume 182, pp. 609–630.
81. Ghadiri, N.; Samani, R.; Shahrokh, F. Integration of Text and Graph-Based Features for Depression Detection Using Visibility Graph. In Proceedings of the 22nd International Conference on Intelligent Systems Design and Applications (ISDA 2022) on Intelligent Systems Design and Applications, Virtual, 12–14 December 2022; Abraham, A., Pllana, S., Casalino, G., Ma, K., Bajaj, A., Eds.; Springer: Cham, Switzerland, 2023; pp. 332–341.
82. Huang, G.; Shen, W.; Lu, H.; Hu, F.; Li, J.; Liu, H. Multimodal Depression Detection based on Factorized Representation. In Proceedings of the 2022 International Conference on High Performance Big Data and Intelligent Systems (HDIS), Tianjin, China, 10–11 December 2022; pp. 190–196. [CrossRef]
83. Liu, D.; Liu, B.; Lin, T.; Liu, G.; Yang, G.; Qi, D.; Qiu, Y.; Lu, Y.; Yuan, Q.; Shuai, S.C.; et al. Measuring depression severity based on facial expression and body movement using deep convolutional neural network. *Front. Psychiatry* **2022**, *13*, 1017064. [CrossRef] [PubMed]
84. Othmani, A.; Zeghina, A.O. A multimodal computer-aided diagnostic system for depression relapse prediction using audiovisual cues: A proof of concept. *Healthc. Anal.* **2022**, *2*, 100090. [CrossRef]
85. Othmani, A.; Zeghina, A.O.; Muzammel, M. A Model of Normality Inspired Deep Learning Framework for Depression Relapse Prediction Using Audiovisual Data. *Comput. Methods Programs Biomed.* **2022**, *226*, 107132. [CrossRef]
86. Park, J.; Moon, N. Design and Implementation of Attention Depression Detection Model Based on Multimodal Analysis. *Sustainability* **2022**, *14*, 3569. [CrossRef]
87. Prabhu, S.; Mittal, H.; Varagani, R.; Jha, S.; Singh, S. Harnessing emotions for depression detection. *Pattern Anal. Appl.* **2022**, *25*, 537–547. [CrossRef]
88. Sudhan, H.V.M.; Kumar, S.S. Multimodal Depression Severity Detection Using Deep Neural Networks and Depression Assessment Scale. In Proceedings of the International Conference on Computational Intelligence and Data Engineering, Vijayawada, India, 12–13 August 2022; Chaki, N., Devarakonda, N., Cortesi, A., Seetha, H., Eds.; Springer: Singapore, 2022; pp. 361–375.
89. T J, S.J.; Jacob, I.J.; Mandava, A.K. D-ResNet-PVKELM: Deep neural network and paragraph vector based kernel extreme machine learning model for multimodal depression analysis. *Multimed. Tools Appl.* **2023**, *82*, 25973–26004. [CrossRef]
90. Vandana; Marriwala, N.; Chaudhary, D. A hybrid model for depression detection using deep learning. *Meas. Sens.* **2023**, *25*, 100587. [CrossRef]

91. Gu, Y.; Zhang, C.; Ma, F.; Jia, X.; Ni, S. AI-Driven Depression Detection Algorithms from Visual and Audio Cues. In Proceedings of the 2023 3rd International Conference on Frontiers of Electronics, Information and Computation Technologies (ICFEICT), Yangzhou, China, 26–29 May 2023; pp. 468–475. [CrossRef]
92. Yoon, J.; Kang, C.; Kim, S.; Han, J. D-vlog: Multimodal Vlog Dataset for Depression Detection. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 12226–12234. [CrossRef]
93. Zhou, L.; Liu, Z.; Shangguan, Z.; Yuan, X.; Li, Y.; Hu, B. TAMFN: Time-Aware Attention Multimodal Fusion Network for Depression Detection. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 669–679. [CrossRef] [PubMed]
94. Zhou, L.; Liu, Z.; Yuan, X.; Shangguan, Z.; Li, Y.; Hu, B. CAIINET: Neural network based on contextual attention and information interaction mechanism for depression detection. *Digit. Signal Process.* **2023**, *137*, 103986. [CrossRef]
95. Qingjun Zhu, J.X.; Peng, L. College students' mental health evaluation model based on tensor fusion network with multimodal data during the COVID-19 pandemic. *Biotechnol. Genet. Eng. Rev.* **2023**, 1–15. [CrossRef]
96. Lam, G.; Dongyan, H.; Lin, W. Context-aware Deep Learning for Multi-modal Depression Detection. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3946–3950. [CrossRef]
97. Niu, M.; Chen, K.; Chen, Q.; Yang, L. HCAG: A Hierarchical Context-Aware Graph Attention Model for Depression Detection. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 4235–4239. [CrossRef]
98. Ma, W.; Qiu, S.; Miao, J.; Li, M.; Tian, Z.; Zhang, B.; Li, W.; Feng, R.; Wang, C.; Cui, Y.; et al. Detecting depression tendency based on deep learning and multi-sources data. *Biomed. Signal Process. Control* **2023**, *86*, 105226. [CrossRef]
99. Thati, R.P.; Dhadwal, A.S.; Kumar, P.; P, S. A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms. *Multimed. Tools Appl.* **2023**, *82*, 4787–4820. [CrossRef] [PubMed]
100. Tlachac, M.; Flores, R.; Reisch, M.; Kayastha, R.; Taurich, N.; Melican, V.; Bruneau, C.; Caouette, H.; Lovering, J.; Toto, E.; et al. StudentSADD: Rapid Mobile Depression and Suicidal Ideation Screening of College Students during the Coronavirus Pandemic. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 1–32. [CrossRef]
101. Su, M.H.; Wu, C.H.; Huang, K.Y.; Yang, T.H. Cell-Coupled Long Short-Term Memory With *L* -Skip Fusion Mechanism for Mood Disorder Detection Through Elicited Audiovisual Features. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 124–135. [CrossRef]
102. Zhang, Z.; Lin, W.; Liu, M.; Mahmoud, M. Multimodal Deep Learning Framework for Mental Disorder Recognition. In Proceedings of the 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Buenos Aires, Argentina, 16–20 November 2020; pp. 344–350. [CrossRef]
103. Ceccarelli, F.; Mahmoud, M. Multimodal temporal machine learning for Bipolar Disorder and Depression Recognition. *Pattern Anal. Appl.* **2022**, *25*, 493–504. [CrossRef]
104. Tlachac, M.; Toto, E.; Lovering, J.; Kayastha, R.; Taurich, N.; Rundensteiner, E. EMU: Early Mental Health Uncovering Framework and Dataset. In Proceedings of the 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Virtual, 13–16 December 2021; pp. 1311–1318. [CrossRef]
105. Tlachac, M.; Flores, R.; Reisch, M.; Houskeeper, K.; Rundensteiner, E.A. DepreST-CAT: Retrospective Smartphone Call and Text Logs Collected during the COVID-19 Pandemic to Screen for Mental Illnesses. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 1–32. [CrossRef]
106. Zhang, Y.; Li, X.; Rong, L.; Tiwari, P. Multi-Task Learning for Jointly Detecting Depression and Emotion. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 3142–3149. [CrossRef]
107. Schultebraucks, K.; Yadav, V.; Shalev, A.Y.; Bonanno, G.A.; Galatzer-Levy, I.R. Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood. *Psychol. Med.* **2022**, *52*, 957–967. [CrossRef] [PubMed]
108. Liu, Y. Using convolutional neural networks for the assessment research of mental health. *Comput. Intell. Neurosci.* **2022**, *2022*, 1636855. [CrossRef] [PubMed]
109. Shen, T.; Jia, J.; Shen, G.; Feng, F.; He, X.; Luan, H.; Tang, J.; Tiropanis, T.; Chua, T.S.; Hall, W. Cross-Domain Depression Detection via Harvesting Social Media. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 1611–1617. [CrossRef]
110. Ricard, B.J.; Marsch, L.A.; Crosier, B.; Hassanpour, S. Exploring the Utility of Community-Generated Social Media Content for Detecting Depression: An Analytical Study on Instagram. *J. Med. Internet Res.* **2018**, *20*, e11817. [CrossRef] [PubMed]
111. Gui, T.; Zhu, L.; Zhang, Q.; Peng, M.; Zhou, X.; Ding, K.; Chen, Z. Cooperative Multimodal Approach to Depression Detection in Twitter. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'19/IAAI'19/EAAI'19), Honolulu, HI, USA, 27 January–1 February 2019; AAAI Press: Washington, DC, USA, 2019. [CrossRef]
112. Wang, Y.; Wang, Z.; Li, C.; Zhang, Y.; Wang, H. A Multimodal Feature Fusion-Based Method for Individual Depression Detection on Sina Weibo. In Proceedings of the 2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC), Austin, TX, USA, 6–8 November 2020; pp. 1–8. [CrossRef]

113. Hu, P.; Lin, C.; Su, H.; Li, S.; Han, X.; Zhang, Y.; Mei, J. BlueMemo: Depression Analysis through Twitter Posts. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence (IJCAI-20), Yokohama, Japan, 11–17 July 2020; pp. 5252–5254. [CrossRef]
114. Li, Y.; Cai, M.; Qin, S.; Lu, X. Depressive Emotion Detection and Behavior Analysis of Men Who Have Sex with Men via Social Media. *Front. Psychiatry* **2020**, *11*, 830. [CrossRef] [PubMed]
115. ALSAGRI, H.S.; YKHLEF, M. Machine Learning-Based Approach for Depression Detection in Twitter Using Content and Activity Features. *IEICE Trans. Inf. Syst.* **2020**, *103*, 1825–1832. [CrossRef]
116. Mann, P.; Paes, A.; Matsushima, E. See and Read: Detecting Depression Symptoms in Higher Education Students Using Multimodal Social Media Data. *Proc. Int. AAAI Conf. Web Soc. Media* **2020**, *14*, 440–451. [CrossRef]
117. Lin, C.; Hu, P.; Su, H.; Li, S.; Mei, J.; Zhou, J.; Leung, H. SenseMood: Depression Detection on Social Media. In Proceedings of the 2020 International Conference on Multimedia Retrieval (ICMR '20), Dublin, Ireland, 8–11 June 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 407–411. [CrossRef]
118. Ghosh, S.; Anwar, T. Depression Intensity Estimation via Social Media: A Deep Learning Approach. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 1465–1474. [CrossRef]
119. Zogan, H.; Razzak, I.; Jameel, S.; Xu, G. DepressionNet: Learning Multi-Modalities with User Post Summarization for Depression Detection on Social Media. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), Paris, France, 21–25 July 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 133–142. [CrossRef]
120. Bi, Y.; Li, B.; Wang, H. Detecting Depression on Sina Microblog Using Depressing Domain Lexicon. In Proceedings of the 2021 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), AB, Canada, 25–28 October 2021; pp. 965–970. [CrossRef]
121. Zhang, Y.; Lyu, H.; Liu, Y.; Zhang, X.; Wang, Y.; Luo, J. Monitoring Depression Trends on Twitter During the COVID-19 Pandemic: Observational Study. *JMIR Infodemiol.* **2021**, *1*, e26769. [CrossRef]
122. Chiu, C.Y.; Lane, H.Y.; Koh, J.L.; Chen, A.L.P. Multimodal depression detection on instagram considering time interval of posts. *J. Intell. Inf. Syst.* **2021**, *56*, 25–47. [CrossRef]
123. Liu, J.; Shi, M. A Hybrid Feature Selection and Ensemble Approach to Identify Depressed Users in Online Social Media. *Front. Psychol.* **2022**, *12*, 802821. [CrossRef]
124. Safa, R.; Bayat, P.; Moghtader, L. Automatic detection of depression symptoms in twitter using multimodal analysis. *J. Supercomput.* **2022**, *78*, 4709–4744. [CrossRef] [PubMed]
125. Cheng, J.C.; Chen, A.L.P. Multimodal time-aware attention networks for depression detection. *J. Intell. Inf. Syst.* **2022**, *59*, 319–339. [CrossRef]
126. Anshul, A.; Pranav, G.S.; Rehman, M.Z.U.; Kumar, N. A Multimodal Framework for Depression Detection During COVID-19 via Harvesting Social Media. *IEEE Trans. Comput. Soc. Syst.* **2023**, 1–17. [CrossRef]
127. Angskun, J.; Tipprasert, S.; Angskun, T. Big data analytics on social networks for real-time depression detection. *J. Big Data* **2022**, *9*, 69. [CrossRef] [PubMed]
128. Uban, A.S.; Chulvi, B.; Rosso, P. Explainability of Depression Detection on Social Media: From Deep Learning Models to Psychological Interpretations and Multimodality. In *Early Detection of Mental Health Disorders by Social Media Monitoring: The First Five Years of the eRisk Project*; Springer: Cham, Switzerland, 2022; pp. 289–320. [CrossRef]
129. Bucur, A.M.; Cosma, A.; Rosso, P.; Dinu, L.P. It's Just a Matter of Time: Detecting Depression with Time-Enriched Multimodal Transformers. In Proceedings of the Advances in Information Retrieval, Dublin, Ireland, 2–6 April 2023; Kamps, J., Goeuriot, L., Crestani, F., Maistro, M., Joho, H., Davis, B., Gurrin, C., Kruschwitz, U., Caputo, A., Eds.; Springer: Cham, Switzerland, 2023; pp. 200–215.
130. Chatterjee, M.; Kumar, P.; Sarkar, D. Generating a Mental Health Curve for Monitoring Depression in Real Time by Incorporating Multimodal Feature Analysis Through Social Media Interactions. *Int. J. Intell. Inf. Technol.* **2023**, *19*, 1–25. [CrossRef]
131. Deng, B.; Wang, Z.; Shu, X.; Shu, J. Transformer-Based Graphic-Text Fusion Depressive Tendency Detection. In Proceedings of the 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 26–29 May 2023; pp. 701–705. [CrossRef]
132. Ghosh, S.; Ekbal, A.; Bhattacharyya, P. What Does Your Bio Say? Inferring Twitter Users' Depression Status From Multimodal Profile Information Using Deep Learning. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 1484–1494. [CrossRef]
133. Jayapal, C.; Yamuna, S.M.; Manavallan, S.; Devasenan, M. Detection of Mental Health Using Deep Learning Technique. In Proceedings of the Communication and Intelligent Systems Dublin, Ireland, 2–6 April 2023; Sharma, H., Shrivastava, V., Bharti, K.K., Wang, L., Eds.; Springer: Singapore, 2023; pp. 507–520.
134. Liaw, A.S.; Chua, H.N. Depression Detection on Social Media With User Network and Engagement Features Using Machine Learning Methods. In Proceedings of the 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 13–15 September 2022; pp. 1–6. [CrossRef]
135. Li, Z.; An, Z.; Cheng, W.; Zhou, J.; Zheng, F.; Hu, B. MHA: A multimodal hierarchical attention model for depression detection in social media. *Health Inf. Sci. Syst.* **2023**, *11*, 6. [CrossRef]

136. Long, X.; Zhang, Y.; Shu, X.; Shu, J. Image-text Fusion Model for Depression Tendency Detection Based on Attention. In Proceedings of the 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 26–29 May 2023; pp. 730–734. [CrossRef]
137. Tong, L.; Liu, Z.; Jiang, Z.; Zhou, F.; Chen, L.; Lyu, J.; Zhang, X.; Zhang, Q.; Sadka, A.; Wang, Y.; et al. Cost-Sensitive Boosting Pruning Trees for Depression Detection on Twitter. *IEEE Trans. Affect. Comput.* **2023**, *14*, 1898–1911. [CrossRef]
138. Pirayesh, J.; Chen, H.; Qin, X.; Ku, W.S.; Yan, D. MentalSpot: Effective Early Screening for Depression Based on Social Contagion. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM '21), Virtual, 1–5 November 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1437–1446. [CrossRef]
139. Mihov, I.; Chen, H.; Qin, X.; Ku, W.S.; Yan, D.; Liu, Y. MentalNet: Heterogeneous Graph Representation for Early Depression Detection. In Proceedings of the 2022 IEEE International Conference on Data Mining (ICDM), Orlando, FL, USA, 28 November–1 December 2022; pp. 1113–1118. [CrossRef]
140. Nuankaew, W.; Doenribram, D.; Jareanpon, C.; Nuankaew, P.; Thanarat, P. A New Probabilistic Weighted Voting Model for Depressive Disorder Classification from Captions and Colors of Images. *ICIC Express Lett.* **2023**, *17*, 531.
141. Suganthi, V.; Punithavalli, M. User Depression and Severity Level Prediction During COVID-19 Epidemic from Social Network Data. *ARPN J. Eng. Appl. Sci.* **2023**, *18*, 1187–1194. [CrossRef]
142. Suri, M.; Semwal, N.; Chaudhary, D.; Gorton, I.; Kumar, B. I Don't Feel so Good! Detecting Depressive Tendencies Using Transformer-Based Multimodal Frameworks. In Proceedings of the 2022 5th International Conference on Machine Learning and Natural Language Processing (MLNLP '22), Xi'an, China, 25–27 March 2022; Association for Computing Machinery: New York, NY, USA, 2023; pp. 360–365. [CrossRef]
143. Valencia-Segura, K.M.; Escalante, H.J.; Villasenor-Pineda, L. Automatic Depression Detection in Social Networks Using Multiple User Characterizations. *Comput. Sist.* **2023**, *27*, 283–294. [CrossRef]
144. Valencia-Segura, K.M.; Escalante, H.J.; Villaseñor-Pineda, L. Leveraging Multiple Characterizations of Social Media Users for Depression Detection Using Data Fusion. In Proceedings of the Mexican Conference on Pattern Recognition, Tepic, Mexico, 21–24 June 2022; Vergara-Villegas, O.O., Cruz-Sánchez, V.G., Sossa-Azuela, J.H., Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., Olvera-López, J.A., Eds.; Springer: Cham, Switzerland, 2022; pp. 215–224.
145. Wang, Y.; Wang, Z.; Li, C.; Zhang, Y.; Wang, H. Online social network individual depression detection using a multitask heterogenous modality fusion approach. *Inf. Sci.* **2022**, *609*, 727–749. [CrossRef]
146. Zogan, H.; Razzak, I.; Wang, X.; Jameel, S.; Xu, G. Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. *World Wide Web* **2022**, *25*, 281–304. [CrossRef] [PubMed]
147. Malhotra, A.; Jindal, R. Multimodal Deep Learning based Framework for Detecting Depression and Suicidal Behaviour by Affective Analysis of Social Media Posts. *EAI Endorsed Trans. Pervasive Health Technol.* **2018**, *6*, 164259. [CrossRef]
148. V, A.M.; C, D.K.A.; S, S.; M, E.; Senthilkumar, M. Cluster Ensemble Method and Convolution Neural Network Model for Predicting Mental Illness. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2023**, *13*, 392–398. [CrossRef]
149. Ghandeharioun, A.; Fedor, S.; Sangermano, L.; Ionescu, D.; Alpert, J.; Dale, C.; Sontag, D.; Picard, R. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 23–26 October 2017; pp. 325–332. [CrossRef]
150. Wang, R.; Wang, W.; daSilva, A.; Huckins, J.F.; Kelley, W.M.; Heatherton, T.F.; Campbell, A.T. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–26. [CrossRef]
151. Xu, X.; Chikersal, P.; Doryab, A.; Villalba, D.K.; Dutcher, J.M.; Tumminia, M.J.; Althoff, T.; Cohen, S.; Creswell, K.G.; Creswell, J.D.; et al. Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1–33. [CrossRef]
152. Masud, M.T.; Rahman, N.; Alam, A.; Griffiths, M.D.; Alamin, M. Non-Pervasive Monitoring of Daily-Life Behavior to Access Depressive Symptom Severity Via Smartphone Technology. In Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 5–7 June 2020; pp. 602–607. [CrossRef]
153. Ware, S.; Yue, C.; Morillo, R.; Lu, J.; Shang, C.; Bi, J.; Kamath, J.; Russell, A.; Bamis, A.; Wang, B. Predicting depressive symptoms using smartphone data. *Smart Health* **2020**, *15*, 100093. [CrossRef]
154. Masud, M.T.; Mamun, M.A.; Thapa, K.; Lee, D.; Griffiths, M.D.; Yang, S.H. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. *J. Biomed. Inform.* **2020**, *103*, 103371. [CrossRef]
155. Chikersal, P.; Doryab, A.; Tumminia, M.; Villalba, D.K.; Dutcher, J.M.; Liu, X.; Cohen, S.; Creswell, K.G.; Mankoff, J.; Creswell, J.D.; et al. Detecting Depression and Predicting Its Onset Using Longitudinal Symptoms Captured by Passive Sensing: A Machine Learning Approach With Robust Feature Selection. *ACM Trans. Comput. Hum. Interact.* **2021**, *28*. [CrossRef]
156. Xu, X.; Chikersal, P.; Dutcher, J.M.; Sefidgar, Y.S.; Seo, W.; Tumminia, M.J.; Villalba, D.K.; Cohen, S.; Creswell, K.G.; Creswell, J.D.; et al. Leveraging Collaborative-Filtering for Personalized Behavior Modeling: A Case Study of Depression Detection among College Students. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–27. [CrossRef]
157. Yan, R.; Liu, X.; Dutcher, J.; Tumminia, M.; Villalba, D.; Cohen, S.; Creswell, D.; Creswell, K.; Mankoff, J.; Dey, A.; et al. A Computational Framework for Modeling Biobehavioral Rhythms from Mobile and Wearable Data Streams. *ACM Trans. Intell. Syst. Technol.* **2022**, *13*, 1–7. [CrossRef]

158. Opoku Asare, K.; Moshe, I.; Terhorst, Y.; Vega, J.; Hosio, S.; Baumeister, H.; Pulkki-Råback, L.; Ferreira, D. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis. *Pervasive Mob. Comput.* **2022**, *83*, 101621. [CrossRef]
159. Suruliraj, B.; Orji, R. Federated Learning Framework for Mobile Sensing Apps in Mental Health. In Proceedings of the 2022 IEEE 10th International Conference on Serious Games and Applications for Health (SeGAH), Sydney, Australia, 10–12 August 2022; pp. 1–7. [CrossRef]
160. Hong, J.; Kim, J.; Kim, S.; Oh, J.; Lee, D.; Lee, S.; Uh, J.; Yoon, J.; Choi, Y. Depressive Symptoms Feature-Based Machine Learning Approach to Predicting Depression Using Smartphone. *Healthcare* **2022**, *10*, 1189. [CrossRef]
161. Kathan, A.; Harrer, M.; Küster, L.; Triantafyllopoulos, A.; He, X.; Milling, M.; Gerczuk, M.; Yan, T.; Rajamani, S.T.; Heber, E.; et al. Personalised depression forecasting using mobile sensor data and ecological momentary assessment. *Front. Digit. Health* **2022**, *4*, 964582. [CrossRef] [PubMed]
162. Kim, J.S.; Wang, B.; Kim, M.; Lee, J.; Kim, H.; Roh, D.; Lee, K.H.; Hong, S.B.; Lim, J.S.; Kim, J.W.; et al. Prediction of Diagnosis and Treatment Response in Adolescents With Depression by Using a Smartphone App and Deep Learning Approaches: Usability Study. *JMIR Form. Res.* **2023**, *7*, e45991. [CrossRef] [PubMed]
163. Liu, Y.; Kang, K.D.; Doe, M.J. HADD: High-Accuracy Detection of Depressed Mood. *Technologies* **2022**, *10*, 123. [CrossRef]
164. Mullick, T.; Radovic, A.; Shaaban, S.; Doryab, A. Predicting Depression in Adolescents Using Mobile and Wearable Sensors: Multimodal Machine Learning–Based Exploratory Study. *JMIR Form. Res.* **2022**, *6*, e35807. [CrossRef]
165. Gerych, W.; Agu, E.; Rundensteiner, E. Classifying Depression in Imbalanced Datasets Using an Autoencoder- Based Anomaly Detection Approach. In Proceedings of the 2019 IEEE 13th International Conference on Semantic Computing (ICSC), Newport Beach, CA, USA, 30 January–1 February 2019; pp. 124–127. [CrossRef]
166. Opoku Asare, K.; Visuri, A.; Vega, J.; Ferreira, D. Me in the Wild: An Exploratory Study Using Smartphones to Detect the Onset of Depression. In Proceedings of the Wireless Mobile Communication and Healthcare, Virtual, 30 November–2 December 2022; Gao, X., Jamalipour, A., Guo, L., Eds.; Springer: Cham, Switzerland, 2022; pp. 121–145.
167. Otte Andersen, T.; Skovlund Dissing, A.; Rosenbek Severinsen, E.; Kryger Jensen, A.; Thanh Pham, V.; Varga, T.V.; Hulvej Rod, N. Predicting stress and depressive symptoms using high-resolution smartphone data and sleep behavior in Danish adults. *Sleep* **2022**, *45*, zsac067. [CrossRef]
168. Tabassum, N.; Ahmed, M.; Shorna, N.J.; Sowad, M.M.U.R.; Haque, H.M.Z. Depression Detection Through Smartphone Sensing: A Federated Learning Approach. *Int. J. Interact. Mob. Technol. (iJIM)* **2023**, *17*, 40–56. [CrossRef]
169. Narziev, N.; Goh, H.; Toshnazarov, K.; Lee, S.A.; Chung, K.M.; Noh, Y. STDD: Short-Term Depression Detection with Passive Sensing. *Sensors* **2020**, *20*, 1396. [CrossRef] [PubMed]
170. Yan, Y.; Tu, M.; Wen, H. A CNN Model with Discretized Mobile Features for Depression Detection. In Proceedings of the 2022 IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks (BSN), Ioannina, Greece, 27–30 September 2022; pp. 1–4. [CrossRef]
171. Zou, B.; Zhang, X.; Xiao, L.; Bai, R.; Li, X.; Liang, H.; Ma, H.; Wang, G. Sequence Modeling of Passive Sensing Data for Treatment Response Prediction in Major Depressive Disorder. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 1786–1795. [CrossRef] [PubMed]
172. Hassantabar, S.; Zhang, J.; Yin, H.; Jha, N.K. MHDeep: Mental Health Disorder Detection System Based on Wearable Sensors and Artificial Neural Networks. *ACM Trans. Embed. Comput. Syst.* **2022**, *21*, 1–22. [CrossRef]
173. Liu, S.; Vahedian, F.; Hachen, D.; Lizardo, O.; Poellabauer, C.; Striegel, A.; Milenković, T. Heterogeneous Network Approach to Predict Individuals' Mental Health. *ACM Trans. Knowl. Discov. Data* **2021**, *15*, 1–26. [CrossRef]
174. Grimm, B.; Talbot, B.; Larsen, L. PHQ-V/GAD-V: Assessments to Identify Signals of Depression and Anxiety from Patient Video Responses. *Appl. Sci.* **2022**, *12*, 9150. [CrossRef]
175. Currey, D.; Torous, J. Digital phenotyping correlations in larger mental health samples: Analysis and replication. *BJPsych Open* **2022**, *8*, e106. [CrossRef]
176. Wang, W.; Nepal, S.; Huckins, J.F.; Hernandez, L.; Vojdanovski, V.; Mack, D.; Plomp, J.; Pillai, A.; Obuchi, M.; daSilva, A.; et al. First-Gen Lens: Assessing Mental Health of First-Generation Students across Their First Year at College Using Mobile Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 1–32. [CrossRef]
177. Thakur, S.S.; Roy, R.B. Predicting mental health using smart-phone usage and sensor data. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 9145–9161. [CrossRef]
178. Choi, J.; Lee, S.; Kim, S.; Kim, D.; Kim, H. Depressed Mood Prediction of Elderly People with a Wearable Band. *Sensors* **2022**, *22*, 4174. [CrossRef]
179. Dai, R.; Kannampallil, T.; Kim, S.; Thornton, V.; Bierut, L.; Lu, C. Detecting Mental Disorders with Wearables: A Large Cohort Study. In Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation (IoTDI '23), San Antonio, TX, USA, 9–12 May 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 39–51. [CrossRef]
180. Dai, R.; Kannampallil, T.; Zhang, J.; Lv, N.; Ma, J.; Lu, C. Multi-Task Learning for Randomized Controlled Trials: A Case Study on Predicting Depression with Wearable Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 1–23. [CrossRef]
181. Horwitz, A.; Czyz, E.; Al-Dajani, N.; Dempsey, W.; Zhao, Z.; Nahum-Shani, I.; Sen, S. Utilizing daily mood diaries and wearable sensor data to predict depression and suicidal ideation among medical interns. *J. Affect. Disord.* **2022**, *313*, 1–7. [CrossRef] [PubMed]

182. Horwitz, A.G.; Kentopp, S.D.; Cleary, J.; Ross, K.; Wu, Z.; Sen, S.; Czyz, E.K. Using machine learning with intensive longitudinal data to predict depression and suicidal ideation among medical interns over time. *Psychol. Med.* **2022**, *53*, 5778–5785. [CrossRef]
183. Shah, A.P.; Vaibhav, V.; Sharma, V.; Al Ismail, M.; Girard, J.; Morency, L.P. Multimodal Behavioral Markers Exploring Suicidal Intent in Social Media Videos. In Proceedings of the 2019 International Conference on Multimodal Interaction (ICMI '19), Suzhou, China, 14–18 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 409–413. [CrossRef]
184. Belouali, A.; Gupta, S.; Sourirajan, V.; Yu, J.; Allen, N.; Alaoui, A.; Dutton, M.A.; Reinhard, M.J. Acoustic and language analysis of speech for suicidal ideation among US veterans. *BioData Min.* **2021**, *14*, 11. [CrossRef] [PubMed]
185. Mishra, R.; Prakhar Sinha, P.; Sawhney, R.; Mahata, D.; Mathur, P.; Ratn Shah, R. SNAP-BATNET: Cascading Author Profiling and Social Network Graphs for Suicide Ideation Detection on Social Media. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop, Dublin, Ireland, 22–27 May 2019; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 147–156. [CrossRef]
186. Ramírez-Cifuentes, D.; Freire, A.; Baeza-Yates, R.; Puntí, J.; Medina-Bravo, P.; Velazquez, D.A.; Gonfaus, J.M.; Gonzàlez, J. Detection of suicidal ideation on social media: Multimodal, relational, and behavioral analysis. *J. Med. Internet Res.* **2020**, *22*, e17758. [CrossRef]
187. Cao, L.; Zhang, H.; Feng, L. Building and Using Personal Knowledge Graph to Improve Suicidal Ideation Detection on Social Media. *IEEE Trans. Multimed.* **2022**, *24*, 87–102. [CrossRef]
188. Chatterjee, M.; Kumar, P.; Samanta, P.; Sarkar, D. Suicide ideation detection from online social media: A multi-modal feature based technique. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100103. [CrossRef]
189. Li, Z.; Cheng, W.; Zhou, J.; An, Z.; Hu, B. Deep learning model with multi-feature fusion and label association for suicide detection. *Multimed. Syst.* **2023**, *29*, 2193–2203. [CrossRef]
190. Heckler, W.F.; Feijó, L.P.; de Carvalho, J.V.; Barbosa, J.L.V. Thoth: An intelligent model for assisting individuals with suicidal ideation. *Expert Syst. Appl.* **2023**, *233*, 120918. [CrossRef]
191. Czyz, E.K.; King, C.A.; Al-Dajani, N.; Zimmermann, L.; Hong, V.; Nahum-Shani, I. Ecological Momentary Assessments and Passive Sensing in the Prediction of Short-Term Suicidal Ideation in Young Adults. *JAMA Netw. Open* **2023**, *6*, e2328005. [CrossRef]
192. Syed, Z.S.; Sidorov, K.; Marshall, D. Automated Screening for Bipolar Disorder from Audio/Visual Modalities. In Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop, Seoul, Republic of Korea, 22 October 2022; Association for Computing Machinery: New York, NY, USA, 2018; pp. 39–45. [CrossRef]
193. Yang, L.; Li, Y.; Chen, H.; Jiang, D.; Oveneke, M.C.; Sahli, H. Bipolar Disorder Recognition with Histogram Features of Arousal and Body Gestures. In Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop (AVEC'18), Seoul, Republic of Korea, 22 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 15–21. [CrossRef]
194. Xing, X.; Cai, B.; Zhao, Y.; Li, S.; He, Z.; Fan, W. Multi-Modality Hierarchical Recall Based on GBDTs for Bipolar Disorder Classification. In Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop (AVEC'18), Seoul, Republic of Korea, 22 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 31–37. [CrossRef]
195. Cao, S.; Yan, H.; Rao, P.; Zhao, K.; Yu, X.; He, J.; Yu, L.; Xiao, Y. Bipolar Disorder Classification Based on Multimodal Recordings. In Proceedings of the 2021 10th International Conference on Computing and Pattern Recognition (ICCPR 2021), Shanghai, China, 15–17 October 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 188–194. [CrossRef]
196. AbaeiKoupaei, N.; Osman, H.A. Multimodal Semi-supervised Bipolar Disorder Classification. In Proceedings of the Intelligent Data Engineering and Automated Learning—IDEAL 2021, Manchester, UK, 25 November 2021; Yin, H., Camacho, D., Tino, P., Allmendinger, R., Tallón-Ballesteros, A.J., Tang, K., Cho, S.B., Novais, P., Nascimento, S., Eds.; Springer: Cham, Switzerland, 2021; pp. 575–586.
197. AbaeiKoupaei, N.; Al Osman, H. A Multi-Modal Stacked Ensemble Model for Bipolar Disorder Classification. *IEEE Trans. Affect. Comput.* **2023**, *14*, 236–244. [CrossRef]
198. Baki, P.; Kaya, H.; Çiftçi, E.; Güleç, H.; Salah, A.A. A Multimodal Approach for Mania Level Prediction in Bipolar Disorder. *IEEE Trans. Affect. Comput.* **2022**, *13*, 2119–2131. [CrossRef]
199. Sivagnanam, L.; Visalakshi, N.K. Multimodal Machine Learning Framework to Detect the Bipolar Disorder. In *Advances in Parallel Computing Algorithms, Tools and Paradigms*; IOS Press: Amsterdam, The Netherlands, 2022. [CrossRef]
200. Su, H.Y.; Wu, C.H.; Liou, C.R.; Lin, E.C.L.; See Chen, P. Assessment of Bipolar Disorder Using Heterogeneous Data of Smartphone-Based Digital Phenotyping. In Proceedings of the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 4260–4264. [CrossRef]
201. Duwairi, R.; Halloush, Z. A Multi-View Learning Approach for Detecting Personality Disorders Among Arab Social Media Users. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2023**, *22*, 1–19. [CrossRef]
202. Bennett, C.C.; Ross, M.K.; Baek, E.; Kim, D.; Leow, A.D. Predicting clinically relevant changes in bipolar disorder outside the clinic walls based on pervasive technology interactions via smartphone typing dynamics. *Pervasive Mob. Comput.* **2022**, *83*, 101598. [CrossRef]
203. Richter, V.; Neumann, M.; Kothare, H.; Roesler, O.; Liscombe, J.; Suendermann-Oeft, D.; Prokop, S.; Khan, A.; Yavorsky, C.; Lindenmayer, J.P.; et al. Towards Multimodal Dialog-Based Speech & Facial Biomarkers of Schizophrenia. In Proceedings of the Companion Publication of the 2022 International Conference on Multimodal Interaction (ICMI '22 Companion), Montreal, QC, Canada, 18–22 October 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 171–176. [CrossRef]

204. Birnbaum, M.L.; Norel, R.; Van Meter, A.; Ali, A.F.; Arenare, E.; Eyigoz, E.; Agurto, C.; Germano, N.; Kane, J.M.; Cecchi, G.A. Identifying signals associated with psychiatric illness utilizing language and images posted to Facebook. *npj Schizophr.* **2020**, *6*, 38. [CrossRef]
205. Wang, R.; Aung, M.S.H.; Abdullah, S.; Brian, R.; Campbell, A.T.; Choudhury, T.; Hauser, M.; Kane, J.; Merrill, M.; Scherer, E.A.; et al. CrossCheck: Toward Passive Sensing and Detection of Mental Health Changes in People with Schizophrenia. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16), Heidelberg, Germany, 12–16 September 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 886–897. [CrossRef]
206. Wang, R.; Wang, W.; Aung, M.S.H.; Ben-Zeev, D.; Brian, R.; Campbell, A.T.; Choudhury, T.; Hauser, M.; Kane, J.; Scherer, E.A.; et al. Predicting Symptom Trajectories of Schizophrenia Using Mobile Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 1–24. [CrossRef]
207. Tseng, V.W.S.; Sano, A.; Ben-Zeev, D.; Brian, R.; Campbell, A.T.; Hauser, M.; Kane, J.M.; Scherer, E.A.; Wang, R.; Wang, W.; et al. Using behavioral rhythms and multi-task learning to predict fine-grained symptoms of schizophrenia. *Sci. Rep.* **2020**, *10*, 15100. [CrossRef]
208. Lamichhane, B.; Zhou, J.; Sano, A. Psychotic Relapse Prediction in Schizophrenia Patients Using A Personalized Mobile Sensing-Based Supervised Deep Learning Model. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 3246–3257. [CrossRef]
209. Zhou, J.; Lamichhane, B.; Ben-Zeev, D.; Campbell, A.; Sano, A. Predicting Psychotic Relapse in Schizophrenia With Mobile Sensor Data: Routine Cluster Analysis. *JMIR mHealth uHealth* **2022**, *10*, e31006. [CrossRef]
210. Osipov, M.; Behzadi, Y.; Kane, J.M.; Petrides, G.; Clifford, G.D. Objective identification and analysis of physiological and behavioral signs of schizophrenia. *J. Ment. Health* **2015**, *24*, 276–282. [CrossRef]
211. Teferra, B.G.; Borwein, S.; DeSouza, D.D.; Rose, J. Screening for Generalized Anxiety Disorder From Acoustic and Linguistic Features of Impromptu Speech: Prediction Model Evaluation Study. *JMIR Form. Res.* **2022**, *6*, e39998. [CrossRef]
212. Choudhary, S.; Thomas, N.; Alshamrani, S.; Srinivasan, G.; Ellenberger, J.; Nawaz, U.; Cohen, R. A Machine Learning Approach for Continuous Mining of Nonidentifiable Smartphone Data to Create a Novel Digital Biomarker Detecting Generalized Anxiety Disorder: Prospective Cohort Study. *JMIR Med. Inform.* **2022**, *10*, e38943. [CrossRef] [PubMed]
213. Ding, Y.; Liu, J.; Zhang, X.; Yang, Z. Dynamic Tracking of State Anxiety via Multi-Modal Data and Machine Learning. *Front. Psychiatry* **2022**, *13*, 757961. [CrossRef] [PubMed]
214. Chen, C.P.; Gau, S.S.F.; Lee, C.C. Learning Converse-Level Multimodal Embedding to Assess Social Deficit Severity for Autism Spectrum Disorder. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 6–10 July 2020; pp. 1–6. [CrossRef]
215. Khullar, V.; Singh, H.P.; Bala, M. Meltdown/Tantrum Detection System for Individuals with Autism Spectrum Disorder. *Appl. Artif. Intell.* **2021**, *35*, 1708–1732. [CrossRef]
216. Mallol-Ragolta, A.; Dhamija, S.; Boult, T.E. A Multimodal Approach for Predicting Changes in PTSD Symptom Severity. In Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18), Boulder, CA, USA, 16–18 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 324–333. [CrossRef]
217. Tébar, B.; Gopalan, A. Early Detection of Eating Disorders using Social Media. In Proceedings of the 2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Orlando, FL, USA, 16–17 December 2021; pp. 193–198. [CrossRef]
218. Abuhassan, M.; Anwar, T.; Liu, C.; Jarman, H.K.; Fuller-Tyszkiewicz, M. EDNet: Attention-Based Multimodal Representation for Classification of Twitter Users Related to Eating Disorders. In Proceedings of the ACM Web Conference 2023 (WWW '23), Houston, TX, USA, 3–6 September 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 4065–4074. [CrossRef]
219. Noguero, D.S.; Ramírez-Cifuentes, D.; Ríssola, E.A.; Freire, A. Gender Bias When Using Artificial Intelligence to Assess Anorexia Nervosa on Social Media: Data-Driven Study. *J. Med. Internet Res.* **2023**, *25*, e45184. [CrossRef] [PubMed]
220. Xu, Z.; Pérez-Rosas, V.; Mihalcea, R. Inferring Social Media Users' Mental Health Status from Multimodal Information. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; European Language Resources Association: Marseille, France, 2020; pp. 6292–6299.
221. Meng, X.; Zhang, J.; Ren, G. The evaluation model of college students' mental health in the environment of independent entrepreneurship using neural network technology. *J. Healthc. Eng.* **2021**, *2021*, 4379623. [CrossRef] [PubMed]
222. Singh, V.K.; Long, T. Automatic assessment of mental health using phone metadata. *Proc. Assoc. Inf. Sci. Technol.* **2018**, *55*, 450–459. [CrossRef]
223. Park, J.; Arunachalam, R.; Silenzio, V.; Singh, V.K. Fairness in Mobile Phone–Based Mental Health Assessment Algorithms: Exploratory Study. *JMIR Form. Res.* **2022**, *6*, e34366. [CrossRef]
224. Liu, S. 3D Illustration of Cartoon Characters Talking And Discussing. Communication and Talking Concept. 3D Rendering on White Background. 2022. Available online: https://www.istockphoto.com/photo/3d-illustration-of-cartoon-characters-talking-and-discussing-communication-and-gm1428415103-471910717 (accessed on 22 November 2023).
225. Arefin, S. Social Media. 2014. Available online: https://www.flickr.com/photos/54888897@N05/5102912860/ (accessed on 10 December 2023).

226. Secret, A. Hand Holding Phone with Social Media Icon Stock Photo. 2021. Available online: https://www.istockphoto.com/photo/hand-holding-phone-with-social-media-icon-gm1351107098-426983736?phrase=smartphone+cartoon (accessed on 10 December 2023).
227. Adventtr. Health Monitoring Information on Generic Smartwatch Screen Stock Photo. 2021. Available online: https://www.istockphoto.com/photo/health-monitoring-information-on-generic-smartwatch-screen-gm1307154121-397513158?utm_source=flickr&utm_medium=affiliate&utm_campaign=srp_photos_top&utm_term=smartphone+and+wearable+cartoon&utm_content=https%3A%2F%2Fwww.flickr.com%2Fsearch%2F&ref=sponsored (accessed on 10 December 2023).
228. Gratch, J.; Artstein, R.; Lucas, G.; Stratou, G.; Scherer, S.; Nazarian, A.; Wood, R.; Boberg, J.; DeVault, D.; Marsella, S.; et al. Th Distress Analysis Interview Corpus of human and computer interviews. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, 26–31 May 2014; European Language Resources Association (ELRA): Reykjavik, Iceland, 2014; pp. 3123–3128.
229. Suendermann-Oeft, D.; Robinson, A.; Cornish, A.; Habberstad, D.; Pautler, D.; Schnelle-Walka, D.; Haller, F.; Liscombe, J.; Neumann, M.; Merrill, M.; et al. NEMSI: A Multimodal Dialog System for Screening of Neurological or Mental Conditions. In Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (IVA '19), Paris, France, 2–5 July 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 245–247. [CrossRef]
230. Çiftçi, E.; Kaya, H.; Güleç, H.; Salah, A.A. The Turkish Audio-Visual Bipolar Disorder Corpus. In Proceedings of the 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), Beijing, China, 20–22 May 2018; pp. 1–6. [CrossRef]
231. Yates, A.; Cohan, A.; Goharian, N. Depression and Self-Harm Risk Assessment in Online Forums. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP), Copenhagen, Denmark, 7—11 September 2017; Association for Computational Linguistics: Copenhagen, Denmark, 2017; pp. 2958–2968.
232. Schueller, S.M.; Begale, M.; Penedo, F.J.; Mohr, D.C. Purple: A Modular System for Developing and Deploying Behavioral Intervention Technologies. *J. Med. Internet Res.* **2014**, *16*, e181. [CrossRef]
233. Farhan, A.A.; Yue, C.; Morillo, R.; Ware, S.; Lu, J.; Bi, J.; Kamath, J.; Russell, A.; Bamis, A.; Wang, B. Behavior vs. introspection: Refining prediction of clinical depression via smartphone sensing data. In Proceedings of the 2016 IEEE Wireless Health (WH), Bethesda, MD, USA, 25–27 October 2016; pp. 1–8. [CrossRef]
234. Montag, C.; Baumeister, H.; Kannen, C.; Sariyska, R.; Meßner, E.M.; Brand, M. Concept, Possibilities and Pilot-Testing of a New Smartphone Application for the Social and Life Sciences to Study Human Behavior Including Validation Data from Personality Psychology. *J* **2019**, *2*, 102–115. [CrossRef]
235. Bai, R.; Xiao, L.; Guo, Y.; Zhu, X.; Li, N.; Wang, Y.; Chen, Q.; Feng, L.; Wang, Y.; Yu, X.; et al. Tracking and Monitoring Mood Stability of Patients With Major Depressive Disorder by Machine Learning Models Using Passive Digital Data: Prospective Naturalistic Multicenter Study. *JMIR Mhealth Uhealth* **2021**, *9*, e24365. [CrossRef]
236. Ferreira, D.; Kostakos, V.; Dey, A.K. AWARE: Mobile Context Instrumentation Framework. *Front. ICT* **2015**, *2*, 6. [CrossRef]
237. Wang, R.; Chen, F.; Chen, Z.; Li, T.; Harari, G.; Tignor, S.; Zhou, X.; Ben-Zeev, D.; Campbell, A.T. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14), Seattle, WA, USA, 13–17 September 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 3–14. [CrossRef]
238. Ringeval, F.; Schuller, B.; Valstar, M.; Gratch, J.; Cowie, R.; Scherer, S.; Mozgai, S.; Cummins, N.; Schmitt, M.; Pantic, M. AVEC 2017: Real-Life Depression, and Affect Recognition Workshop and Challenge. In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (AVEC '17), Mountain View, CA, USA, 23–27 October 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 3–9. [CrossRef]
239. Ringeval, F.; Schuller, B.; Valstar, M.; Cummins, N.; Cowie, R.; Tavabi, L.; Schmitt, M.; Alisamir, S.; Amiriparian, S.; Messner, E.M.; et al. AVEC 2019 Workshop and Challenge: State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19), Nice, France, 21–25 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 3–12. [CrossRef]
240. Dhamija, S.; Boult, T.E. Exploring Contextual Engagement for Trauma Recovery. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 2267–2277. [CrossRef]
241. Orton, I. Vision based body gesture meta features for Affective Computing. *arXiv* **2020**, *arXiv:2003.00809*.
242. Cohan, A.; Desmet, B.; Yates, A.; Soldaini, L.; MacAvaney, S.; Goharian, N. SMHD: A Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions. In Proceedings of the 27th International Conference on Computational Linguistics (COLING), Santa Fe, NM, USA, 20–26 August 2018; Association for Computational Linguistics: Dublin, Ireland, 2018; pp. 1485–1497.
243. Cao, L.; Zhang, H.; Feng, L.; Wei, Z.; Wang, X.; Li, N.; He, X. Latent Suicide Risk Detection on Microblog via Suicide-Oriented Word Embeddings and Layered Attention. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Bali, Indonesia, 8–12 December 2019; Association for Computational Linguistics: Dublin, Ireland, 2019; pp. 1718–1728. [CrossRef]
244. Wang, X.; Chen, S.; Li, T.; Li, W.; Zhou, Y.; Zheng, J.; Zhang, Y.; Tang, B. Assessing depression risk in Chinese microblogs: A corpus and machine learning methods. In Proceedings of the 2019 IEEE International Conference on Healthcare Informatics (ICHI), Xi'an, China, 10–13 June 2019; pp. 1–5. [CrossRef]

245. Losada, D.E.; Crestani, F. A Test Collection for Research on Depression and Language Use. In Proceedings of the 7th International Conference of the Cross-Language Evaluation Forum for European Languages, Evora, Portugal, 5–8 September 2016; Experimental IR Meets Multilinguality, Multimodality, and Interaction; Springer: Cham, Switzerland, 2016; pp. 28–39.
246. Losada, D.E.; Crestani, F.; Parapar, J. Overview of eRisk: Early Risk Prediction on the Internet. In Proceedings of the Experimental IR Meets Multilinguality, Multimodality, and Interaction, Avignon, France, 10–14 September 2018; Bellot, P., Trabelsi, C., Mothe, J., Murtagh, F., Nie, J.Y., Soulier, L., SanJuan, E., Cappellato, L., Ferro, N., Eds.; Springer: Cham, Switzerland, 2018; pp. 343–361.
247. Vesel, C.; Rashidisabet, H.; Zulueta, J.; Stange, J.P.; Duffecy, J.; Hussain, F.; Piscitello, A.; Bark, J.; Langenecker, S.A.; Young, S.; et al. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: A BiAffect iOS study. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1007–1018. [CrossRef] [PubMed]
248. Mattingly, S.M.; Gregg, J.M.; Audia, P.; Bayraktaroglu, A.E.; Campbell, A.T.; Chawla, N.V.; Das Swain, V.; De Choudhury, M.; D'Mello, S.K.; Dey, A.K.; et al. The Tesserae Project: Large-Scale, Longitudinal, In Situ, Multimodal Sensing of Information Workers. In Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19), Glasgow, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–8. [CrossRef]
249. Coppersmith, G.; Dredze, M.; Harman, C.; Hollingshead, K.; Mitchell, M. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Denver CO, USA, 31 July 2015; Association for Computational Linguistics: Denver, CO, USA, 2015; pp. 31–39. [CrossRef]
250. Denny, J.C.; Rutter, J.L.; Goldstein, D.B.; Philippakis, A.; Smoller, J.W.; Jenkins, G.; Dishman, E. The "All of Us" Research Program. *New Engl. J. Med.* **2019**, *381*, 668–676. [CrossRef] [PubMed]
251. Ramírez-Cifuentes, D.; Freire, A.; Baeza-Yates, R.; Lamora, N.S.; Álvarez, A.; González-Rodríguez, A.; Rochel, M.L.; Vives, R.L.; Velazquez, D.A.; Gonfaus, J.M.; et al. Characterization of Anorexia Nervosa on Social Media: Textual, Visual, Relational, Behavioral, and Demographical Analysis. *J. Med. Internet Res.* **2021**, *23*, e25925. [CrossRef] [PubMed]
252. Teferra, B.G.; Borwein, S.; DeSouza, D.D.; Simpson, W.; Rheault, L.; Rose, J. Acoustic and Linguistic Features of Impromptu Speech and Their Association With Anxiety: Validation Study. *JMIR Ment. Health* **2022**, *9*, e36828. [CrossRef]
253. Palan, S.; Schitter, C. Prolific.ac—A subject pool for online experiments. *J. Behav. Exp. Financ.* **2018**, *17*, 22–27. [CrossRef]
254. Hamilton, M. A Rating Scale for Depression. *J. Neurol. Neurosurg. Psychiatry* **1960**, *23*, 56–62. [CrossRef]
255. Kroenke, K.; Spitzer, R.L. The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatr. Ann.* **2002**, *32*, 509–515. [CrossRef]
256. Beck, A.T.; Ward, C.H.; Mendelson, M.; Mock, J.; Erbaugh, J. An Inventory for Measuring Depression. *Arch. Gen. Psychiatry* **1961**, *4*, 561–571. [CrossRef]
257. Radloff, L.S. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Appl. Psychol. Meas.* **1977**, *1*, 385–401. [CrossRef]
258. Kroenke, K.; Spitzer, R.L.; Williams, J.B. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* **2001**, *16*, 606–613. [CrossRef] [PubMed]
259. Aytar, Y.; Vondrick, C.; Torralba, A. SoundNet: Learning Sound Representations from Unlabeled Video. In Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16), Barcelona, Spain, 5–10 December 2016; Curran Associates Inc.: Red Hook, NY, USA, 2016; pp. 892–900.
260. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
261. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 2nd International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014; pp. 892–900.
262. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [CrossRef]
263. Ekman, P. Basic Emotions. In *Handbook of Cognition and Emotion*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 1999; Chapter 3; pp. 45–60. [CrossRef]
264. Plutchik, R. Chapter 1—A General Psychoevolutionary Theory of Emotion. In *Theories of Emotion*; Plutchik, R., Kellerman, H., Eds.; Academic Press: Cambridge, MA, USA,1980; pp. 3–33. [CrossRef]
265. Perronnin, F.; Sánchez, J.; Mensink, T. Improving the Fisher Kernel for Large-Scale Image Classification. In Proceedings of the Computer Vision (ECCV 2010), Heraklion, Greece, 5–11 September 2010; Daniilidis, K., Maragos, P., Paragios, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 143–156.
266. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 4171–4186. [CrossRef]
267. Eyben, F.; Wöllmer, M.; Schuller, B. Opensmile: The Munich Versatile and Fast Open-Source Audio Feature Extractor. In Proceedings of the 18th ACM International Conference on Multimedia (MM '10), Firenze, Italy, 25–29 October 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 1459–1462. [CrossRef]
268. Crocco, M.; Cristani, M.; Trucco, A.; Murino, V. Audio Surveillance: A Systematic Review. *ACM Comput. Surv.* **2016**, *48*, 1–46. [CrossRef]

269. Baltrusaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.P. OpenFace 2.0: Facial Behavior Analysis Toolkit. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018. [CrossRef]
270. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* **2000**, *25*, 120–123.
271. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
272. Ekman, P.; Friesen, W.V. *Facial Action Coding System*; Consulting Psychologists Press: Washington, DC, USA, 1978. [CrossRef]
273. Prince, E.B.; Martin, K.B.; Messinger, D.S. Facial Action Coding System. In *The SAGE Encyclopedia of Communication Research Methods*; SAGE Publications, Inc.: London, UK, 2017. [CrossRef]
274. Zhi, R.; Liu, M.; Zhang, D. A comprehensive survey on automatic facial action unit analysis. *Vis. Comput.* **2020**, *36*, 1067–1093. [CrossRef]
275. Lin, C.; Mottaghi, S.; Shams, L. The effects of color and saturation on the enjoyment of real-life images. *Psychon. Bull. Rev.* **2023**, *30*, 1–12. [CrossRef]
276. Valdez, P.; Mehrabian, A. Effects of color on emotions. *J. Exp. Psychol. Gen.* **1994**, *123*, 394–409. [CrossRef]
277. Hashemipour, S.; Ali, M. Amazon Web Services (AWS)—An Overview of the On-Demand Cloud Computing Platform. In Proceedings of the Emerging Technologies in Computing, Virtual, 27–29 August 2020; Miraz, M.H., Excell, P.S., Ware, A., Soomro, S., Ali, M., Eds.; Springer: Cham, Switzerland, 2020; pp. 40–47.
278. Pennebaker Conglomerates, Inc. Linguistic Inquiry and Word Count: LIWC-22. 2022. Available online: https://www.liwc.app (accessed on 10 December 2023).
279. NLP Tools for the Social Sciences. Suite of Automatic Linguistic Analysis Tools (SALAT). 2023. Available online: https://www.linguisticanalysistools.org/ (accessed on 10 December 2023).
280. Bird, S.; Loper, E. NLTK: The Natural Language Toolkit. In Proceedings of the ACL Interactive Poster and Demonstration Sessions, Stroudsburg, PA, USA, 21–26 July 2004; Association for Computational Linguistics: Barcelona, Spain, 2004; pp. 214–217.
281. Manning, C.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S.; McClosky, D. The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, MD, USA, 22–27 June 2014; Association for Computational Linguistics: Baltimore, MD, USA, 2014; pp. 55–60. [CrossRef]
282. Crossley, S.A.; Kyle, K.; McNamara, D.S. Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behav. Res. Methods* **2017**, *49*, 803–821. [CrossRef]
283. Bradley, M.M.; Lang, P.J. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*; Technical Report C-1; The Center for Research in Psychophysiology: University of Florida, Gainesville, FL, USA, 1999.
284. Le, Q.; Mikolov, T. Distributed Representations of Sentences and Documents. In Proceedings of the 31st International Conference on International Conference on Machine Learning (ICML'14), Stockholm, Sweden, 10–15 July 2014; Volume 32, pp. II–1188–II–1196.
285. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.; Le, Q.V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Curran Associates Inc.: Red Hook, NY, USA, 2019.
286. Hakulinen, C.; Elovainio, M.; Pulkki-Råback, L.; Virtanen, M.; Kivimäki, M.; Jokela, M. Personality and depressive symptoms: Individual participant meta-analysis of 10 cohort studies. *Depress. Anxiety* **2015**, *32*, 461–470. [CrossRef]
287. Greenspon, T.S. Is there an Antidote to Perfectionism? *Psychol. Sch.* **2014**, *51*, 986–998. [CrossRef]
288. Clark-Carter, D. z Scores. In *Wiley StatsRef: Statistics Reference Online*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2014. [CrossRef]
289. Mahesh, B. Machine Learning Algorithms—A Review. *Int. J. Sci. Res.* **2020**, *9*, 381–386.
290. Wang, S.C. Artificial Neural Network. In *Interdisciplinary Computing in Java Programming*; Springer: Boston, MA, USA, 2003; pp. 81–100. [CrossRef]
291. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]
292. Sagi, O.; Rokach, L. Ensemble learning: A survey. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]
293. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]
294. Graves, A. Supervised Sequence Labelling With Recurrent Neural Networks. In *Studies in Computational Intelligence;* Springer: Berlin/Heidelberg, Germany, 2012; Volume 385. [CrossRef]
295. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 3–8 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 6000–6010.
296. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692.
297. Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. In Proceedings of the 8th International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
298. Kim, T.; Vossen, P. EmoBERTa: Speaker-Aware Emotion Recognition in Conversation with RoBERTa. *arXiv* **2021**, arXiv:2108.12009.

299. Rasheed, K.; Qayyum, A.; Ghaly, M.; Al-Fuqaha, A.; Razi, A.; Qadir, J. Explainable, trustworthy, and ethical machine learning for healthcare: A survey. *Comput. Biol. Med.* **2022**, *149*, 106043. [CrossRef]
300. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
301. Freund, Y.; Schapire, R.E. A desicion-theoretic generalization of on-line learning and an application to boosting. In Proceedings of the Computational Learning Theory, Barcelona, Spain, 13–15 March 1995; Vitányi, P., Ed.; Springer: Berlin/Heidelberg, Germany, 1995; pp. 23–37.
302. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
303. Lui, M. Feature Stacking for Sentence Classification in Evidence-Based Medicine. In Proceedings of the Australasian Language Technology Association Workshop 2012, Dunedin, New Zealand, 4–6 December 2012; pp. 134–138.
304. Xu, S.; An, X.; Qiao, X.; Zhu, L.; Li, L. Multi-output least-squares support vector regression machines. *Pattern Recognit. Lett.* **2013**, *34*, 1078–1084. [CrossRef]
305. Rasmus, A.; Berglund, M.; Honkala, M.; Valpola, H.; Raiko, T. Semi-supervised Learning with Ladder Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2015; Volume 28.
306. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
307. Ricci, F.; Rokach, L.; Shapira, B. Introduction to Recommender Systems Handbook. In *Recommender Systems Handbook*; Springer: Boston, MA, USA, 2011; pp. 1–35. [CrossRef]
308. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
309. Tang, J.; Wang, K. Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM ’18), Los Angeles, CA, USA, 15–29 September 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 565–573. [CrossRef]
310. Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* **2021**, *54*, 1–35. [CrossRef]
311. Amiri, Z.; Heidari, A.; Darbandi, M.; Yazdani, Y.; Jafari Navimipour, N.; Esmaeilpour, M.; Sheykhi, F.; Unal, M. The Personal Health Applications of Machine Learning Techniques in the Internet of Behaviors. *Sustainability* **2023**, *15*, 2406. [CrossRef]
312. Adler, D.A.; Wang, F.; Mohr, D.C.; Choudhury, T. Machine learning for passive mental health symptom prediction: Generalization across different longitudinal mobile sensing studies. *PLoS ONE* **2022**, *17*, e0266516. [CrossRef]
313. Morgan, C.; Tonkin, E.L.; Craddock, I.; Whone, A.L. Acceptability of an In-home Multimodal Sensor Platform for Parkinson Disease: Nonrandomized Qualitative Study. *JMIR Hum. Factors* **2022**, *9*, e36370. [CrossRef]
314. McCarney, R.; Warner, J.; Iliffe, S.; van Haselen, R.; Griffin, M.; Fisher, P. The Hawthorne Effect: A randomised, controlled trial. *BMC Med. Res. Methodol.* **2007**, *7*, 30. [CrossRef]
315. American Psychiatric Publishing. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5™*, 5th ed.; American Psychiatric Publishing: Washington, DC, USA, 2013.
316. Hussain, M.; Al-Haiqi, A.; Zaidan, A.; Zaidan, B.; Kiah, M.; Anuar, N.B.; Abdulnabi, M. The landscape of research on smartphone medical apps: Coherent taxonomy, motivations, open challenges and recommendations. *Comput. Methods Programs Biomed.* **2015**, *122*, 393–408. [CrossRef]
317. Tsai, J.; Kelley, P.; Cranor, L.; Sadeh, N. Location-Sharing Technologies: Privacy Risks and Controls. *Innov. Law Policy eJournal* **2009**, *6*, 119.
318. Taylor, J.; Pagliari, C. Mining social media data: How are research sponsors and researchers addressing the ethical challenges? *Res. Ethics* **2018**, *14*, 1–39. [CrossRef]
319. Mavrogiorgou, A.; Kleftakis, S.; Mavrogiorgos, K.; Zafeiropoulos, N.; Menychtas, A.; Kiourtis, A.; Maglogiannis, I.; Kyriazis, D. beHEALTHIER: A Microservices Platform for Analyzing and Exploiting Healthcare Data. In Proceedings of the 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Aveiro, Portugal, 7–9 June 2021; pp. 283–288. [CrossRef]
320. Georgogiannis, A.; Digalakis, V. Speech Emotion Recognition using non-linear Teager energy based features in noisy environments. In Proceedings of the 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, Romania, 27–31 August 2012; pp. 2045–2049.
321. Degottex, G.; Kane, J.; Drugman, T.; Raitio, T.; Scherer, S. COVAREP—A collaborative voice analysis repository for speech technologies. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 960–964.
322. Mathieu, B.; Essid, S.; Fillon, T.; Prado, J.; Richard, G. YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software. In Proceedings of the 11th International Society for Music Information Retrieval Conference, Utrecht, The Netherlands, 9–13 August 2010; pp. 441–446. Available online: http://ismir2010.ismir.net/proceedings/ismir2010-75.pdf (accessed on 10 December 2023).
323. Jadoul, Y.; Thompson, B.; de Boer, B. Introducing Parselmouth: A Python interface to Praat. *J. Phon.* **2018**, *71*, 1–15. [CrossRef]
324. Giannakopoulos, T. pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. *PLoS ONE* **2015**, *10*, e0144610. [CrossRef] [PubMed]

325. Orozco-Arroyave, J.R.; Vásquez-Correa, J.C.; Vargas-Bonilla, J.F.; Arora, R.; Dehak, N.; Nidadavolu, P.; Christensen, H.; Rudzicz, F.; Yancheva, M.; Chinaei, H.; et al. NeuroSpeech: An open-source software for Parkinson's speech analysis. *Digit. Signal Process.* **2018**, *77*, 207–221. [CrossRef]
326. MYOLUTION Lab My-Voice-Analysis. 2018. Available online: https://github.com/Shahabks/my-voice-analysis (accessed on 10 December 2023).
327. Lenain, R.; Weston, J.; Shivkumar, A.; Fristed, E. Surfboard: Audio Feature Extraction for Modern Machine Learning. In Proceedings of the 21th Annual Conference of the International Speech Communication Association (INTERSPEECH 2020), Shanghai, China, 25–29 October 2020; pp. 2917–2921. [CrossRef]
328. McFee, B.; Raffel, C.; Liang, D.; Ellis, D.P.W.; McVicar, M.; Battenberg, E.; Nieto, O. librosa: Audio and Music Signal Analysis in Python. In Proceedings of the 14th Python in Science Conference, Austin, TX, USA, 6–12 July 2015; pp. 18–24. [CrossRef]
329. Schuller, B.; Steidl, S.; Batliner, A.; Burkhardt, F.; Devillers, L.; Müller, C.; Narayanan, S. The INTERSPEECH 2010 paralinguistic challenge. In Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010), Makuhari, Japan, 26–30 September 2010; pp. 2794–2797. [CrossRef]
330. Schuller, B.; Steidl, S.; Batliner, A.; Vinciarelli, A.; Scherer, K.; Ringeval, F.; Chetouani, M.; Weninger, F.; Eyben, F.; Marchi, E.; et al. The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. In Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH 2013), Lyon, France, 25–29 August 2013; pp. 148–152. [CrossRef]
331. Eyben, F.; Scherer, K.R.; Schuller, B.W.; Sundberg, J.; André, E.; Busso, C.; Devillers, L.Y.; Epps, J.; Laukka, P.; Narayanan, S.S.; et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Trans. Affect. Comput.* **2016**, *7*, 190–202. [CrossRef]
332. Hannun, A.; Case, C.; Casper, J.; Catanzaro, B.; Diamos, G.; Elsen, E.; Prenger, R.; Satheesh, S.; Sengupta, S.; Coates, A.; et al. Deep Speech: Scaling up end-to-end speech recognition. *arXiv* **2014**, arXiv:1412.5567. [CrossRef]
333. Hershey, S.; Chaudhuri, S.; Ellis, D.P.W.; Gemmeke, J.F.; Jansen, A.; Moore, R.C.; Plakal, M.; Platt, D.; Saurous, R.A.; Seybold, B.; et al. CNN architectures for large-scale audio classification. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–19 March 2017; pp. 131–135. [CrossRef]
334. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]
335. Ravanelli, M.; Bengio, Y. Interpretable Convolutional Filters with SincNet. *arXiv* **2018**, arXiv:1811.09725.
336. Baevski, A.; Zhou, Y.; Mohamed, A.; Auli, M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 12449–12460.
337. Lin, Z.; Feng, M.; dos Santos, C.N.; Yu, M.; Xiang, B.; Zhou, B.; Bengio, Y. A Structured Self-Attentive Sentence Embedding. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
338. Hsu, W.N.; Bolte, B.; Tsai, Y.H.H.; Lakhotia, K.; Salakhutdinov, R.; Mohamed, A. HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **2021**, *29*, 3451–3460. [CrossRef]
339. Huang, Z.; Zhang, J.; Ma, L.; Mao, F. GTCN: Dynamic Network Embedding Based on Graph Temporal Convolution Neural Network. In Proceedings of the Intelligent Computing Theories and Application, Bari, Italy, 2–5 October 2020; Huang, D.S., Jo, K.H., Eds.; Springer: Cham, Switzerland, 2020; pp. 583–593.
340. Schmitt, M.; Schuller, B. openXBOW—Introducing the Passau Open-Source Crossmodal Bag-of-Words Toolkit. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
341. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555. [CrossRef]
342. Viola, P.; Jones, M. Robust Real-Time Object Detection. *Int. J. Comput. Vis. IJCV* **2001**, *57*, 5385–5395.
343. Tzimiropoulos, G.; Pantic, M. Gauss-Newton Deformable Part Models for Face Alignment In-the-Wild. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1851–1858. [CrossRef]
344. Jeni, L.A.; Cohn, J.F.; Kanade, T. Dense 3D face alignment from 2D videos in real-time. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 1.
345. Zhou, E.; Fan, H.; Cao, Z.; Jiang, Y.; Yin, Q. Extensive Facial Landmark Localization with Coarse-to-Fine Convolutional Network Cascade. In Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops, Sydney, Australia, 2–8 December 2013; pp. 386–391. [CrossRef]
346. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17), San Francisco, CA, USA, 4–9 February 2017; AAAI Press: Washington, DC, USA, 2017; pp. 4278–4284.
347. King, D.E. Dlib-ml: A Machine Learning Toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.
348. Tzimiropoulos, G.; Alabort-i Medina, J.; Zafeiriou, S.; Pantic, M. Generic Active Appearance Models Revisited. In Proceedings of the Computer Vision (ACCV 2012), Daejeon, Republic of Korea, 5–9 November 2012; Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 650–663.

349. Zhou, B.; Lapedriza, A.; Khosla, A.; Oliva, A.; Torralba, A. Places: A 10 Million Image Database for Scene Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1452–1464. [CrossRef] [PubMed]
350. Onal Ertugrul, I.; Jeni, L.A.; Ding, W.; Cohn, J.F. AFAR: A Deep Learning Based Tool for Automated Facial Affect Recognition. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019.
351. Meng, H.; Pears, N.; Bailey, C. A Human Action Recognition System for Embedded Computer Vision Application. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–6. [CrossRef]
352. Face++ AI Open Platform. Face++. 2012. Available online: https://www.faceplusplus.com/ (accessed on 20 September 2023).
353. Littlewort, G.; Whitehill, J.; Wu, T.; Fasel, I.; Frank, M.; Movellan, J.; Bartlett, M. The computer expression recognition toolbox (CERT). In Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, 21–25 March 2011; pp. 298–305. [CrossRef]
354. Meng, H.; Huang, D.; Wang, H.; Yang, H.; AI-Shuraifi, M.; Wang, Y. Depression Recognition Based on Dynamic Facial and Vocal Expression Features Using Partial Least Square Regression. In Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '13), Barcelona, Spain, 21 October 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 21–30. [CrossRef]
355. Mower, E.; Matarić, M.J.; Narayanan, S. A Framework for Automatic Human Emotion Classification Using Emotion Profiles. *IEEE Trans. Audio Speech Lang. Process.* **2011**, *19*, 1057–1070. [CrossRef]
356. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5987–5995. [CrossRef]
357. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 7–10 September 2015; Xie, X., Jones, M.W., Tam, G.K.L., Eds.; BMVA Press: Durham, UK, 2015; pp. 41.1–41.12. [CrossRef]
358. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
359. Kollias, D.; Tzirakis, P.; Nicolaou, M.A.; Papaioannou, A.; Zhao, G.; Schuller, B.; Kotsia, I.; Zafeiriou, S. Deep Affect Prediction in-the-Wild: Aff-Wild Database and Challenge, Deep Architectures, and Beyond. *Int. J. Comput. Vis.* **2019**, *127*, 907–929. [CrossRef]
360. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning (PMLR), Long Beach, CA, USA, 9–15 June 2019. [CrossRef]
361. Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; Joulin, A. Emerging Properties in Self-Supervised Vision Transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 9650–9660.
362. Hoffstaetter, S.; Bochi, J.; Lee, M.; Kistner, L.; Mitchell, R.; Cecchini, E.; Hagen, J.; Morawiec, D.; Bedada, E.; Akyüz, U. Pytesseract: A Python wrapper for Google Tesseract. 2019. Available online: https://github.com/madmaze/pytesseract (accessed on 20 September 2023).
363. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision. In Proceedings of the 38th International Conference on Machine Learning (PMLR, 2021), Virtual, 18–24 July 2021; Volume 139, pp. 8748–8763.
364. Technologies Imagga. Imagga. 2023. Available online: https://imagga.com/ (accessed on 20 September 2023).
365. Sikka, K.; Wu, T.; Susskind, J.; Bartlett, M. Exploring Bag of Words Architectures in the Facial Expression Domain. In Proceedings of the Computer Vision—ECCV 2012. Workshops and Demonstrations, Florence, Italy, 7–13 October 2012; Fusiello, A., Murino, V., Cucchiara, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 250–259.
366. van de Weijer, J.; Schmid, C.; Verbeek, J.; Larlus, D. Learning Color Names for Real-World Applications. *IEEE Trans. Image Process.* **2009**, *18*, 1512–1523. [CrossRef]
367. Lin, H.; Jia, J.; Guo, Q.; Xue, Y.; Li, Q.; Huang, J.; Cai, L.; Feng, L. User-Level Psychological Stress Detection from Social Media Using Deep Neural Network. In Proceedings of the 22nd ACM International Conference on Multimedia (MM '14), Orlando, FL, USA, 3–7 November 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 507–516. [CrossRef]
368. Ibraheem, N.; Hasan, M.; Khan, R.Z.; Mishra, P. Understanding Color Models: A Review. *ARPN J. Sci. Technol.* **2012**, *2*, 265–275.
369. Ramírez-esparza, N.; Pennebaker, J.W.; Andrea García, F.; Amd suriá, R. La psicología del uso de las palabras: Un programa de computadora que analiza textos en español. *Rev. Mex. Psicol.* **2007**, *24*, 85–99
370. Lv, M.; Li, A.; Liu, T.; Zhu, T. Creating a Chinese suicide dictionary for identifying suicide risk on social media. *PeerJ* **2015**, *3*, e1455. [CrossRef]
371. Huang, C.L.; Chung, C.; Hui, N.; Lin, Y.C.; Seih, Y.T.; Lam, B.; Pennebaker, J. Development of the Chinese linguistic inquiry and word count dictionary. *Chin. J. Psychol.* **2012**, *54*, 185–201.
372. Gao, R.; Hao, B.; Li, H.; Gao, Y.; Zhu, T. Developing Simplified Chinese Psychological Linguistic Analysis Dictionary for Microblog. In Proceedings of the Brain and Health Informatics: International Conference (BHI 2013), Maebashi, Japan, 29–31 October 2013; Imamura, K., Usui, S., Shirao, T., Kasamatsu, T., Schwabe, L., Zhong, N., Eds.; Springer: Cham, Switzerland, 2013; pp. 359–368.

373. Crossley, S.A.; Allen, L.K.; Kyle, K.; McNamara, D.S. Analyzing Discourse Processing Using a Simple Natural Language Processing Tool. *Discourse Process.* **2014**, *51*, 511–534. [CrossRef]
374. Das Swain, V.; Chen, V.; Mishra, S.; Mattingly, S.M.; Abowd, G.D.; De Choudhury, M. Semantic Gap in Predicting Mental Wellbeing through Passive Sensing. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), New Orleans, LA, USA, 3–5 May 2022; Association for Computing Machinery: New York, NY, USA, 2022. [CrossRef]
375. Sun, J. *Jieba Chinese Word Segmentation Tool*; ACM: New York, NY, USA, 2012.
376. Loria, S.; Keen, P.; Honnibal, M.; Yankovsky, R.; Karesh, D.; Dempsey, E.; Childs, W.; Schnurr, J.; Qalieh, A.; Ragnarsson, L.; et al. TextBlob: Simplified Text Processing. 2013. Available online: https://textblob.readthedocs.io/en/dev/ (accessed on 10 December 2023).
377. Marcus, M.P.; Santorini, B.; Marcinkiewicz, M.A. Building a Large Annotated Corpus of English: The Penn Treebank. *Comput. Linguist.* **1993**, *19*, 313–330.
378. Fast, E.; Chen, B.; Bernstein, M.S. Empath. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems ACM, San Jose, CA, USA, 7–12 May 2016. [CrossRef]
379. Zubiaga, A. TF-CR: Weighting Embeddings for Text Classification. *arXiv* **2020**, arXiv:2012.06606. [CrossRef]
380. Bansal, S.; Aggarwal, C. *Textstat*; Freie University Berlin: Berlin, Germany, 2014.
381. Wenliang, C.; Jingbo, Z.; Muhua, Z.; Tianshun, Y. Text Representation Using Domain Dictionary. *J. Comput. Res. Dev.* **2005**, *42*, 2155.
382. Li, G.; Li, B.; Huang, L.; Hou, S. Automatic Construction of a Depression-Domain Lexicon Based on Microblogs: Text Mining Study. *JMIR Med. Inform.* **2020**, *8*, e17650. [CrossRef]
383. Mohammad, S.M.; Turney, P.D. *NRC Emotion Lexicon*; Technical Report; National Research Council of Canada: Montreal, QC, Canada, 2013. [CrossRef]
384. Hofman, E. *Senti-py: A Sentiment Analysis Classifier in Spanish*; Springer: Berlin/Heidelberg, Germany, 2018.
385. Hutto, C.; Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proc. Int. Aaai Conf. Web Soc. Media* **2014**, *8*, 216–225. [CrossRef]
386. School of Computer Science and Technology. *Chinese Emotion Lexicons*; School of Computer Science and Technology:Luton, UK, 2020.
387. Mohammad, S.M.; Turney, P.D. Crowdsourcing a Word–Emotion Association Lexicon. *Comput. Intell.* **2013**, *29*, 436–465. [CrossRef]
388. Cambria, E.; Speer, R.; Havasi, C.; Hussain, A. SenticNet: A Publicly Available Semantic Resource for Opinion Mining. In Proceedings of the AAAI Fall Symposium: Commonsense Knowledge, Arlington, VA, USA, 11–13 November 2010.
389. Namenwirth, J. *The Lasswell Value Dictionary*; Springer: Berlin/Heidelberg, Germany, 1968.
390. Nielsen, F.Å. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings, Heraklion, Crete, 30 May 2011; pp. 93–98.
391. Dodds, P.S.; Harris, K.D.; Kloumann, I.M.; Bliss, C.A.; Danforth, C.M. Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. *PLoS ONE* **2011**, *6*, e026752. [CrossRef]
392. Gupta, A.; Band, A.; Sharma, S.R. text2emotion. 2020. Available online: https://github.com/aman2656/text2emotion-library (accessed on 20 September 2023).
393. Kralj Novak, P.; Smailović, J.; Sluban, B.; Mozetič, I. Sentiment of Emojis. *PLoS ONE* **2015**, *10*, e.0144296. [CrossRef]
394. Ren, F.; Kang, X.; Quan, C. Examining Accumulated Emotional Traits in Suicide Blogs With an Emotion Topic Model. *IEEE J. Biomed. Health Inform.* **2016**, *20*, 1384–1396. [CrossRef] [PubMed]
395. Pennington, J.; Socher, R.; Manning, C.D. GloVe: Global Vectors for Word Representation. In Proceedings of the Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
396. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed Representations of Words and Phrases and Their Compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13), Lake Tahoe, NV, USA, 3–6 December 2013; Curran Associates Inc.: Red Hook, NY, USA, 2013; Volume 2, pp. 3111–3119.
397. Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. Enriching Word Vectors with Subword Information. *Trans. Assoc. Comput. Linguist.* **2016**, *5*, 135–146. [CrossRef]
398. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep Contextualized Word Representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), New Orleans, LA, USA, 1–6 June 2018; Association for Computational Linguistics: New Orleans, LA, USA, 2018; pp. 2227–2237. [CrossRef]
399. Deriu, J.; Lucchi, A.; De Luca, V.; Severyn, A.; Müller, S.; Cieliebak, M.; Hofmann, T.; Jaggi, M. Leveraging Large Amounts of Weakly Supervised Data for Multi-Language Sentiment Classification. In Proceedings of the 26th International Conference on World Wide Web (CHE, 2017; WWW '17), Perth, Australia, 3–7 May 2017; pp. 1045–1052. [CrossRef]
400. Wang, W.; Wei, F.; Dong, L.; Bao, H.; Yang, N.; Zhou, M. MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 6–12 December 2020; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 5776–5788.

401. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 6–12 December 2020; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 1877–1901.
402. Alshubaily, I. TextCNN with Attention for Text Classification. *arXiv* **2021**, arXiv:2108.01921.
403. Cer, D.; Yang, Y.; Kong, S.y.; Hua, N.; Limtiaco, N.; John, R.S.; Constant, N.; Guajardo-Cespedes, M.; Yuan, S.; Tar, C.; et al. Universal Sentence Encoder. *arXiv* **2018**, arXiv:1803.11175. [CrossRef]
404. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv* **2019**, arXiv:1908.10084.
405. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
406. Yan, X.; Guo, J.; Lan, Y.; Cheng, X. A Biterm Topic Model for Short Texts. In Proceedings of the Proceedings of the 22nd International Conference on World Wide Web (WWW '13), Virtual, 13–17 May 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 1445–1456. [CrossRef]
407. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Virtual, 5–10 July 2020; Association for Computational Linguistics: Toronto, ON, Canada, 2020; pp. 7871–7880. [CrossRef]
408. Wang, J.; Song, Y.; Leung, T.; Rosenberg, C.; Wang, J.; Philbin, J.; Chen, B.; Wu, Y. Learning Fine-Grained Image Similarity with Deep Ranking. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2014; pp. 1386–1393. [CrossRef]
409. Bromley, J.; Bentz, J.W.; Bottou, L.; Guyon, I.M.; LeCun, Y.; Moore, C.; Säckinger, E.; Shah, R. Signature Verification Using A "Siamese" Time Delay Neural Network. *Int. J. Pattern Recognit. Artif. Intell.* **1993**, *7*, 669–688. [CrossRef]
410. Li, Q.; Xue, Y.; Zhao, L.; Jia, J.; Feng, L. Analyzing and Identifying Teens' Stressful Periods and Stressor Events From a Microblog. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 1434–1448. [CrossRef]
411. Grover, A.; Leskovec, J. Node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 855–864. [CrossRef]
412. Dunbar, R.; Arnaboldi, V.; Conti, M.; Passarella, A. The structure of online social networks mirrors those in the offline world. *Soc. Netw.* **2015**, *43*, 39–47. [CrossRef]
413. Vega, J.; Li, M.; Aguillera, K.; Goel, N.; Joshi, E.; Khandekar, K.; Durica, K.C.; Kunta, A.R.; Low, C.A. Reproducible Analysis Pipeline for Data Streams: Open-Source Software to Process Data Collected With Mobile Devices. *Front. Digit. Health* **2021**, *3*, 769823. [CrossRef]
414. Sak, H.; Senior, A.W.; Beaufays, F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In Proceedings of the Interspeech, Singapore, 14–18 September 2014.
415. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96), Portland, OR, USA, 2–4 August 1996; AAAI Press: Washington, DC, USA, 1996; pp. 226–231.
416. Saeb, S.; Zhang, M.; Karr, C.J.; Schueller, S.M.; Corden, M.E.; Kording, K.P.; Mohr, D.C. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *J. Med. Internet Res.* **2015**, *17*, e175. [CrossRef] [PubMed]
417. Canzian, L.; Musolesi, M. Trajectories of Depression: Unobtrusive Monitoring of Depressive States by Means of Smartphone Mobility Traces Analysis. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15), Osaka, Japan, 7–11 September 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 1293–1304. [CrossRef]
418. Cornelissen, G. Cosinor-based rhythmometry. *Theor. Biol. Med. Model.* **2014**, *11*, 16. [CrossRef] [PubMed]
419. Barandas, M.; Folgado, D.; Fernandes, L.; Santos, S.; Abreu, M.; Bota, P.; Liu, H.; Schultz, T.; Gamboa, H. TSFEL: Time Series Feature Extraction Library. *SoftwareX* **2020**, *11*, 100456. [CrossRef]
420. Geary, D.N.; McLachlan, G.J.; Basford, K.E. Mixture Models: Inference and Applications to Clustering. *J. R. Stat. Soc. Ser. (Statistics Soc.)* **1989**, *152*, 126. [CrossRef]
421. Kaufmann, L.; Rousseeuw, P. Clustering by Means of Medoids. *Data Analysis Based on the L1-Norm and Related Methods*; KU Leuven: Leuven, Belgium, 1987; pp. 405–416.
422. Shi, C.; Li, Y.; Zhang, J.; Sun, Y.; Yu, P.S. A survey of heterogeneous information network analysis. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 17–37. [CrossRef]
423. Farseev, A.; Samborskii, I.; Chua, T.S. BBridge: A Big Data Platform for Social Multimedia Analytics. In Proceedings of the 24th ACM International Conference on Multimedia, (MM '16), Vancouver, BC, Canada, 26–31 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 759–761. [CrossRef]
424. Sap, M.; Park, G.; Eichstaedt, J.C.; Kern, M.L.; Stillwell, D.J.; Kosinski, M.; Ungar, L.H.; Schwartz, H.A. Developing age and gender predictive lexica over social media. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014.

425. Wang, Z.; Hale, S.; Adelani, D.I.; Grabowicz, P.; Hartman, T.; Flöck, F.; Jurgens, D. Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. In Proceedings of the The World Wide Web Conference ACM, Amsterdam, The Netherlands, 21–23 October 2019. [CrossRef]
426. The International Business Machines Corporation (IBM). IBM Watson Natural Language Understanding. 2021. Available online: https://www.ibm.com/products/natural-language-understanding (accessed on 20 September 2023).
427. Mehta, Y.; Fatehi, S.; Kazameini, A.; Stachl, C.; Cambria, E.; Eetemadi, S. Bottom-Up and Top-Down: Predicting Personality with Psycholinguistic and Language Model Features. In Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 17–20 November 2020; pp. 1184–1189. [CrossRef]
428. Sun, B.; Li, L.; Wu, X.; Zuo, T.; Chen, Y.; Zhou, G.; He, J.; Zhu, X. Combining feature-level and decision-level fusion in a hierarchical classifier for emotion recognition in the wild. *J. Multimodal User Interfaces* **2016**, *10*, 125–137. [CrossRef]
429. Arevalo, J.; Solorio, T.; Montes-y Gómez, M.; González, F.A. Gated multimodal networks. *Neural Comput. Appl.* **2020**, *32*, 10209–10228. [CrossRef]
430. Kim, J.H.; On, K.W.; Lim, W.; Kim, J.; Ha, J.W.; Zhang, B.T. Hadamard Product for Low-rank Bilinear Pooling. *arXiv* **2016**, arXiv:1610.04325. [CrossRef]
431. Fukui, A.; Park, D.H.; Yang, D.; Rohrbach, A.; Darrell, T.; Rohrbach, M. Multimodal Compact Bilinear Pooling for Visual Question Answering and Visual Grounding. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Austin, TX, USA, 1–5 November 2016. [CrossRef]
432. Liu, Z.; Shen, Y.; Lakshminarasimhan, V.B.; Liang, P.P.; Zadeh, A.B.; Morency, L.P. Efficient Low-rank Multimodal Fusion With Modality-Specific Factors. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Toronto, ON, Canada, 2018. [CrossRef]
433. Yu, Z.; Yu, J.; Fan, J.; Tao, D. Multi-modal Factorized Bilinear Pooling with Co-attention Learning for Visual Question Answering. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [CrossRef]
434. Tan, H.; Bansal, M. LXMERT: Learning Cross-Modality Encoder Representations from Transformers. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Qingdao, China, 13–17 October 2019; Association for Computational Linguistics: Hong Kong, China, 2019; pp. 5100–5111. [CrossRef]
435. Zadeh, A.; Liang, P.P.; Mazumder, N.; Poria, S.; Cambria, E.; Morency, L.P. Memory Fusion Network for Multi-View Sequential Learning. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'18/IAAI'18/EAAI'18), New Orleans, LA, USA, 2–7 February 2018; AAAI Press: Washington, DC, USA, 2018.
436. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. (Methodol.)* **1996**, *58*, 267–288. [CrossRef]
437. Zou, H.; Hastie, T.J. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. (Statistical Methodol.)* **2005**, *67*, 301–320. [CrossRef]
438. de Jong, S. SIMPLS: An alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 251–263. [CrossRef]
439. Schölkopf, B.; Platt, J.C.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C. Estimating the Support of a High-Dimensional Distribution. *Neural Comput.* **2001**, *13*, 1443–1471. [CrossRef] [PubMed]
440. Fokkema, M.; Smits, N.; Zeileis, A.; Hothorn, T.; Kelderman, H. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behav. Res. Methods* **2017**, *50*, 2016–2034. [CrossRef] [PubMed]
441. van den Oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. WaveNet: A Generative Model for Raw Audio. *arXiv* **2016**, arXiv:1609.03499.
442. Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; Xu, B. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), East Stroudsburg, PA, USA, 7–12 August 2016; Association for Computational Linguistics: Berlin, Germany, 2016; pp. 207–212. [CrossRef]
443. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [CrossRef]
444. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical Attention Networks for Document Classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; Association for Computational Linguistics: San Diego, CA, USA, 2016; pp. 1480–1489. [CrossRef]
445. Chen, C.; Breiman, L. *Using Random Forest to Learn Imbalanced Data*; University of California: Berkeley, CA, USA, 2004.
446. Zong, W.; Huang, G.B.; Chen, Y. Weighted extreme learning machine for imbalance learning. *Neurocomputing* **2013**, *101*, 229–242. [CrossRef]
447. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, Hungary, 25–29 July 2004; Volume 2, pp. 985–990. [CrossRef]
448. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for Activation Functions. *arXiv* **2018**, arXiv:1710.05941.

449. Pezeshki, M.; Fan, L.; Brakel, P.; Courville, A.; Bengio, Y. Deconstructing the Ladder Network Architecture. In Proceedings of the 33rd International Conference on Machine Learning (PMLR), New York, NY, USA, 20–22 June 2016; Volume 48, pp. 2368–2376.
450. Drumond, L.R.; Diaz-Aviles, E.; Schmidt-Thieme, L.; Nejdl, W. Optimizing Multi-Relational Factorization Models for Multiple Target Relations. In Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM '14), Shanghai, China, 3–7 November 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 191–200. [CrossRef]
451. Nickel, M.; Tresp, V.; Kriegel, H.P. A Three-Way Model for Collective Learning on Multi-Relational Data. In Proceedings of the 28th International Conference on Machine Learning (ICML'11), Bellevue, WA, USA, 28 June–2 July 2011; Omnipress: Madison, WI, USA, 2011; pp. 809–816.
452. Bader, B.W.; Harshman, R.A.; Kolda, T.G. Temporal Analysis of Semantic Graphs Using ASALSAN. In Proceedings of the Seventh IEEE International Conference on Data Mining (ICDM 2007), Omaha, NE, USA, 28–31 October 2007; pp. 33–42. [CrossRef]
453. Shi, C.; Hu, B.; Zhao, W.; Yu, P. Heterogeneous Information Network Embedding for Recommendation. *IEEE Trans. Knowl. Data Eng.* **2017**, *31*, 357–370. [CrossRef]
454. Milenković, T.; Przulj, N. Uncovering biological network function via graphlet degree signatures. *Cancer Inform.* **2008**, *6*, 257–273. [CrossRef]
455. Gu, S.; Johnson, J.; Faisal, F.E.; Milenković, T. From homogeneous to heterogeneous network alignment via colored graphlets. *Sci. Rep.* **2017**, *8*, 12524. [CrossRef]
456. Perozzi, B.; Al-Rfou, R.; Skiena, S. DeepWalk. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM, New York, NY, USA, 24–27 August 2014. [CrossRef]
457. Dong, Y.; Chawla, N.V.; Swami, A. Metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17), Halifax, NS, USA, 13–17 August 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 135–144. [CrossRef]
458. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation-Based Anomaly Detection. *ACM Trans. Knowl. Discov. Data* **2012**, *6*, 1–39. [CrossRef]
459. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying Density-Based Local Outliers. *SIGMOD Rec.* **2000**, *29*, 93–104. [CrossRef]
460. Feasel, K. Connectivity-Based Outlier Factor (COF). In *Finding Ghosts in Your Data: Anomaly Detection Techniques with Examples in Python*; Apress: Berkeley, CA, USA, 2022; pp. 185–201. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.