A Survey of Depression Detection and Assessment Using Multimodal and Unimodal Methods

Abstract

This survey paper provides a comprehensive analysis of current methodologies in depression detection and assessment, emphasizing the integration of advanced technologies such as deep learning and neural networks. Depression, a prevalent mental health disorder, poses significant challenges in diagnosis due to the limitations of traditional methods, which rely heavily on subjective measures. The paper highlights the transformative potential of leveraging multimodal and unimodal approaches, integrating diverse data types such as audio, visual, and text cues, to enhance detection accuracy and reliability. The review explores the role of emotion recognition and affective computing in understanding psychological abnormalities, underscoring the importance of facial and vocal emotion recognition in mental health assessments. It also discusses the advancements in deep learning models, showcasing their applications in improving the detection and understanding of depression. Key challenges, including data quality, model generalization, and interpretability, are addressed, along with privacy and ethical considerations in deploying these technologies. The paper concludes by emphasizing the need for ongoing research and development to enhance the efficacy of depression detection technologies, ultimately improving mental health outcomes and intervention strategies.

1 Introduction

1.1 Global Impact of Depression

Depression is a significant global health concern, affecting over 264 million individuals and recognized as a leading cause of disability [1]. Approximately 3.8% of the global population experiences depression-related disorders, characterized by persistent low mood, diminished interest in activities, and substantial impairments across personal, social, academic, and occupational domains [2, 3]. Untreated depression can escalate to suicidal ideation, emphasizing its profound impact on individual well-being [4].

The COVID-19 pandemic has exacerbated mental health challenges, leading to increased depression rates and highlighting the urgent need for early detection and intervention. This crisis has intensified the economic and societal burdens associated with depression, necessitating innovative detection and treatment strategies to mitigate its effects [5].

Technological advancements in the digital era provide new opportunities for enhancing mental well-being. There is growing interest in employing machine learning and other advanced technologies to improve mental health outcomes, offering personalized and effective interventions to address the escalating global mental health crisis [6]. The rising prevalence of depression necessitates ongoing research and development to tackle this pressing health challenge.

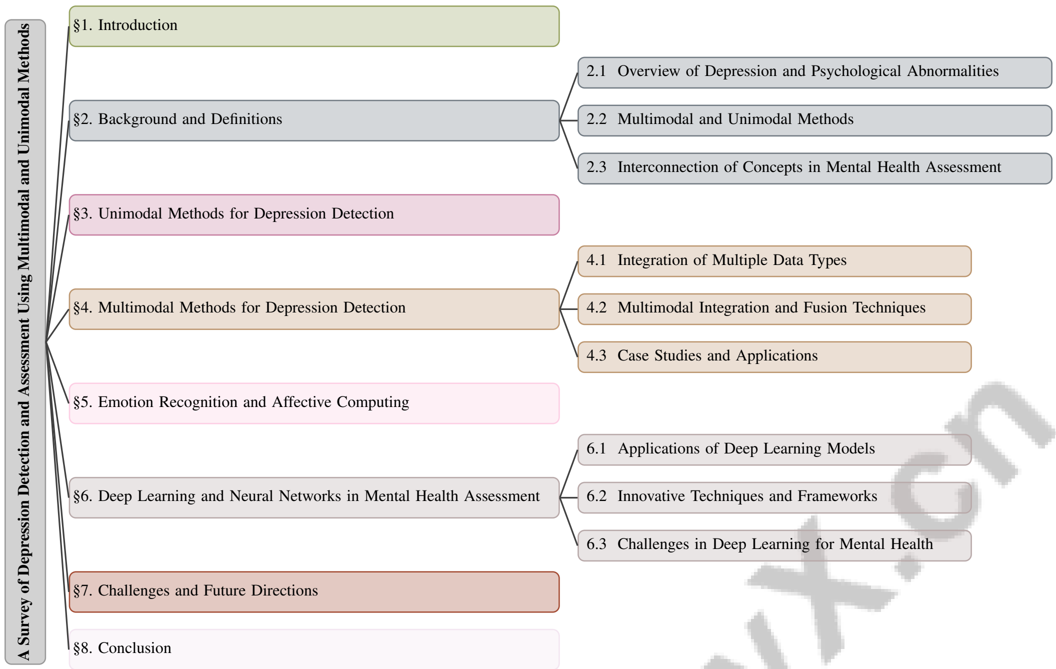


Figure 1: chapter structure

1.2 Limitations of Traditional Methods

Traditional diagnostic methods for depression, including clinical interviews and self-reported questionnaires, face significant challenges due to their reliance on subjective measures, which can result in misdiagnosis and delayed treatment. These methods often lack objective biophysical evidence, leading to assessments that may not capture the complexity of Major Depressive Disorder (MDD) [7]. Compounding these limitations are the adverse effects of mental illness on functioning, low mental health literacy, and poor help-seeking attitudes, underscoring the necessity for innovative detection strategies [8].

Existing diagnostic approaches primarily differentiate between depressed and non-depressed individuals, lacking the nuance required to classify varying levels of depression severity [1]. In primary healthcare settings, approximately 25% of depression diagnoses by physicians are inaccurate [9]. Furthermore, traditional methods often overlook the involvement of family members, which could enhance understanding of the patient's condition [10]. The reluctance of adolescents to express negative emotions and seek psychological help further illustrates the inadequacy of these methods in effectively supporting younger populations [11].

Challenges also arise in detecting and interpreting linguistic nuances indicative of depression, particularly in social media contexts, where traditional methods struggle to identify these cues. Current techniques often rely on single-dimensional feature models, potentially overlooking the wealth of information embedded in diverse speech characteristics [3]. Additionally, traditional methods are labor-intensive and time-consuming, limiting their overall effectiveness [12].

Despite advancements in artificial intelligence and natural language processing, significant gaps remain in the automated identification of emotions related to mental disorders, especially in regions where stigma hinders help-seeking behaviors [13]. Limitations in contextual understanding within lexicon-based approaches and the opaque nature of deep learning models in social media analysis further highlight the shortcomings of existing machine learning methods [14]. These challenges underscore the urgent need for innovative, automated systems that can integrate diverse data sources while ensuring transparency and interpretability in mental health assessments.

1.3 Advancements in Technology

Technological advancements have revolutionized depression detection and assessment, introducing methodologies that enhance precision and efficiency beyond traditional approaches. The integration of artificial intelligence (AI) and deep learning has facilitated the development of advanced systems capable of analyzing diverse data sources, including text, audio, video, and physiological signals, to detect mood disorders. Automatic Depression Detection (ADD) systems have emerged as promising solutions, although challenges persist in accurately extracting depression-specific information from lengthy video sequences [15]. These systems utilize convolutional neural networks (CNNs) and multipart interactive training to improve speech-based depression screening [16].

The rise of social media provides a unique opportunity to analyze individuals' mental health through their online expressions, crucial for understanding and diagnosing depression. Advances in deep learning have led to smartphone applications that assess preliminary depression status by analyzing social media data, showcasing the potential of digital platforms in mental health evaluation [11]. The implementation of psychiatric scale-guided risky post screening further enhances the efficiency and accuracy of depression detection by filtering relevant posts.

Multimodal approaches, which integrate various data types, have shown promise in improving depression detection accuracy. The growing interest in multimodal machine learning within precision health, particularly for clinical decision support through data fusion, highlights the need for timely diagnosis [17]. Although existing methods have not adequately addressed this need, the fusion of audio-visual information using deep learning techniques has improved detection efficiency, despite high costs associated with equipment like fMRI and EEG [18].

AI chatbots have emerged as valuable tools for mental health diagnostics, particularly for depressive disorders, offering an accessible alternative for emotional expression and psychological counseling [11]. Virtual agents present an economical option for gathering and analyzing clinical interview data, making the application of machine and deep learning in depression detection an important research area [19]. The introduction of the InterMind system, which includes family members in the diagnostic and treatment processes, exemplifies the potential of interactive depression assessment systems [10].

A systematic review of affective computing highlights the integration of emotion recognition and sentiment analysis, addressing knowledge gaps in existing literature by covering emotion models, databases, and recent advancements [20]. Researchers are employing deep learning to identify patterns associated with depression, illustrating the transformative impact of technology in this field [21]. Moreover, topic modeling-based methods have been proposed to segment interviews into topics, enabling context-aware feature extraction from audio, video, and text, thus enhancing the understanding of depression-related cues [22]. These advancements reflect the ongoing evolution of tools and methodologies in mental health assessment, offering new pathways for personalized and effective mental health interventions.

1.4 Structure of the Survey

This survey is meticulously structured to provide a comprehensive analysis of current methodologies in depression detection and assessment, emphasizing the integration of advanced technologies. The paper begins with an **Introduction**, which sets the context by discussing the global impact of depression and the limitations of traditional diagnostic methods. This section also highlights recent technological advancements that have transformed mental health assessment, paving the way for innovative detection strategies.

Following the introduction, the **Background and Definitions** section offers an overview of depression and related psychological abnormalities. It defines critical concepts such as multimodal and unimodal methods, emotion recognition, affective computing, deep learning, and neural networks, explaining their interconnections in mental health assessment.

The survey then delves into **Unimodal Methods for Depression Detection**, exploring approaches that utilize a single type of data, such as text, audio, or video. This section provides a detailed examination of text-based, audio-based, and video-based methods, including facial expression analysis, and discusses the role of social media in detecting depression symptoms.

Next, the paper explores **Multimodal Methods for Depression Detection**, emphasizing the benefits of integrating multiple data types to enhance detection accuracy. This section discusses techniques for data integration and presents case studies demonstrating the effectiveness of multimodal approaches.

The role of **Emotion Recognition and Affective Computing** is then examined, highlighting how these technologies contribute to understanding and assessing depression. This section discusses methods for recognizing emotions through facial and vocal cues, the importance of linguistic and acoustic features, and the applications of emotion recognition in real-world scenarios.

The survey continues with a focus on **Deep Learning and Neural Networks in Mental Health Assessment**, analyzing their application in improving the detection and understanding of depression. The paper provides an in-depth examination of cutting-edge models and innovative techniques in the realm of mental health analytics, particularly focusing on the detection of mental disorders through social media platforms. It highlights the latest advancements in machine learning and explainable AI (XAI) methodologies, while also addressing significant challenges such as the non-standard language found in user-generated content and the ethical considerations surrounding data collection and processing. Additionally, the discussion encompasses various multimodal approaches that integrate textual, audio, and visual data to enhance the accuracy of mental health assessments [23, 24, 25, 20, 26].

In the **Challenges and Future Directions** section, the paper identifies current challenges, such as data privacy and model generalization, and discusses potential future directions and research opportunities to address these issues. It also examines data quality and availability, model interpretability, and ethical considerations in deploying advanced technologies for mental health.

The **Conclusion** synthesizes the primary findings of the survey, highlighting the critical role of advanced technologies in enhancing mental health assessments. It underscores the potential of these technologies to significantly improve the detection and treatment of depression by leveraging multimodal data analysis, which incorporates visual, textual, and interactional cues from social media. This integration not only facilitates more accurate identification of depressive behaviors but also enables the development of demographic-aware health interventions, ultimately leading to more effective early detection and personalized treatment strategies for individuals at risk of depression [27, 28, 29, 26, 30].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Overview of Depression and Psychological Abnormalities

Major Depressive Disorder (MDD) is a pervasive mental health issue marked by persistent sadness, anhedonia, cognitive impairments, and disruptions in daily activities [31]. It significantly affects personal and social spheres, potentially leading to severe outcomes like suicide. Symptoms include guilt, low self-worth, disturbed sleep or appetite, low energy, and poor concentration. Diagnosing depression is challenging due to its subjective nature and the high false positive rates of traditional self-reported screening methods [5], underscoring the need for more reliable diagnostic tools.

Depression often coexists with conditions like anxiety and bipolar disorder, necessitating continuous symptom tracking through clinical interviews [32]. Fine-grained classification techniques are crucial for distinguishing between overlapping symptoms, particularly in text-based analyses [24]. Advances in machine learning and natural language processing offer promising enhancements in detection, especially for patients with comorbid conditions. The COVID-19 pandemic has exacerbated depression prevalence, highlighting its societal and individual impact. Innovative computational methods that analyze atypical emotions are vital for early detection and timely interventions [6].

The lack of explainability in current deep learning models, particularly on social media, undermines trust among health professionals. Addressing this is crucial for developing reliable diagnostic tools capable of accurate depression diagnosis through interactive conversations. Integrating speech emotion recognition into mental health systems can significantly enhance early detection and treatment, especially amid rising pandemic-induced mental health issues. Advanced deep learning models analyzing vocal characteristics and emotional cues can enhance voice-enabled devices, aiding in identifying conditions like depression and schizophrenia through multimodal speech analysis [33, 34, 35, 36]. These developments emphasize the need for interdisciplinary approaches to capture the dynamic nature of depression and its associated conditions.

Challenges in obtaining sufficient sample data for effective depression detection, particularly in neuroimaging (MRI, fMRI) and EEG-based studies, can lead to overfitting [37]. This highlights the importance of comprehensive datasets and innovative methods. Exploring bodily gestures as indicators of psychological distress offers a less-explored avenue compared to vocal and facial modalities, enriching multimodal approaches to depression detection.

2.2 Multimodal and Unimodal Methods

In mental health assessment, distinguishing between unimodal and multimodal methods is crucial for detecting depression and related abnormalities. Unimodal methods focus on a single data type, such as text, audio, or visual cues, offering simplicity and resource efficiency [31]. Textual analysis, for instance, has been used to classify tweets as depressive or non-depressive, illustrating a unimodal approach [2]. However, unimodal methods may fall short in capturing the complex nature of mental health conditions, as transient emotions, especially in audio data, challenge the detection of stable conditions like depression [38]. Vision-based body gesture meta-features, which aggregate gesture features for predicting psychological distress, highlight the potential of visual data [19].

Conversely, multimodal methods integrate various data types—audio, text, and visual cues—to provide a comprehensive analysis of an individual's mental state. By leveraging multiple modalities, these approaches enhance the accuracy and reliability of assessments [2]. Techniques like the Attention Fusion Network exemplify this integration by combining video, text, and audio data for improved depression detection [22]. Multimodal methods address the challenge of accurately detecting depression from diverse sources.

Multimodal approaches are categorized into early, intermediate, and late fusion techniques, each offering different integration strategies. For instance, combining speech analysis with cognitive tasks, such as the n-Back working memory task, illustrates a multimodal approach to isolating core depression symptoms [6]. Challenges persist in integrating temporal information and addressing the 'black box' nature of AI models, which often lack interpretability [38].

While unimodal methods provide foundational insights, multimodal approaches offer significant advantages in accuracy and depth. The development of innovative techniques and datasets continues to advance depression detection and assessment. The complexity of integrating modalities like ECG and EEG poses challenges due to the lack of standardized datasets for AI model training [19]. Employing Skeleton Data Augmentation enhances dataset quality, crucial for improving detection accuracy. These advancements reflect the continuous evolution of methodologies in mental health assessment.

2.3 Interconnection of Concepts in Mental Health Assessment

Assessing mental health, particularly depression, involves a complex interplay of concepts and methodologies utilizing unimodal and multimodal approaches. These methods aim to enhance the reliability and interpretability of detection by grounding it in clinically relevant symptoms. Integrating emotional expressions, such as those related to speech and facial features, serves as indicators of mental disorders [39]. This integration is essential for differentiating between conditions like dementia and depression, underscoring the need for improved diagnostic tools [40].

Advanced machine learning techniques, such as Gate Recurrent Unit (GRU) and Bidirectional Long Short-Term Memory (BiLSTM) models with attention layers, process audio and text features for more accurate depression detection [41]. These models illustrate the potential of combining different data types for a comprehensive understanding of a patient's mental health [42]. The fusion of speech and linguistic representations further enhances detection accuracy, showcasing the effectiveness of multimodal approaches [43].

The creation of multimodal datasets, such as those combining EEG and audio data from clinically diagnosed patients, provides benchmarks for evaluating depression detection models [44]. These datasets are crucial for measuring models' ability to classify depression levels accurately, contributing to robust assessment frameworks [45]. Categorizing existing research into various frameworks emphasizes integrating multiple modalities to enhance assessments [46].

Incorporating questionnaire-based mental health assessments into chatbots represents an innovative approach to establishing automated systems for diagnostics [47]. This trend reflects the broader movement of utilizing technology for comprehensive and accessible assessments. Ongoing advancements in these interconnected concepts highlight the potential for developing more effective tools, ultimately improving outcomes for individuals facing mental health challenges.

Integrating articulatory coordination features from speech with contextual information from text enhances depression classification accuracy, illustrating the synergy between modalities [48]. A sub-attention mechanism linking diverse modalities further improves estimation accuracy [49]. Machine learning and natural language processing techniques offer promising opportunities for screening, monitoring, early detection, and prevention of adverse outcomes in mental disorders [50]. Additionally, the cognitive network framework aids in categorizing emotional expressions in texts, revealing the complexity and duality of feelings expressed, vital for understanding assessments [51].

The integration of chain-of-thought (CoT) prompting into AI models has been proposed to enhance reasoning processes when determining PHQ-8 scores, improving interpretability and accuracy [29]. Utilizing neural networks to analyze linguistic features allows effective identification of psychological states, particularly in depression [52]. A multimodal deep learning framework integrating audio and text features enhances mental disorder prediction, reflecting the multimodal nature of clinical decision-making. The framework for understanding emotion recognition through unimodal and multimodal approaches categorizes existing research based on modalities, providing a comprehensive overview of the field [20].

3 Unimodal Methods for Depression Detection

Category	Feature	Method
Text-Based Approaches	Emotion Integration Multimodal Approaches Task-Oriented Learning Data-Driven Learning	HAN+L[53] VBGMF[54] MT-BGRU[55] SSL-DD[56]
Audio-Based Approaches	Feature Integration Techniques Remote Assessment Data	ASSD[16] ADRA[9]
Video and Facial Expression Analysis	Temporal Dynamics Analysis Feature Fusion Techniques	MD-CNN-LSTM[1] ABAFnet[3]
Social Media and Online Interactions	Multimodal Analysis Data Summarization Sentiment Analysis	CD[57], MDD[58] DN[59] DeSK[60]

Table 1: This table provides a comprehensive overview of unimodal methods for depression detection, categorizing them into text-based, audio-based, video and facial expression analysis, and social media interactions. Each category is further subdivided into specific techniques, highlighting the diverse methodologies and advancements in mental health assessment. The table serves as a valuable resource for understanding the application and integration of various approaches in detecting depressive symptoms.

Table 5 provides a detailed classification of unimodal methods for depression detection, showcasing the various data types and techniques utilized in text, audio, and video-based approaches. Unimodal methods, characterized by the use of a single data type, play a crucial role in the detection of depression by focusing on specific patterns within text, audio, or visual data. This section delves into various unimodal approaches, beginning with text-based methods that utilize linguistic analysis to uncover indicators of mental health disorders. Table 1 presents a detailed classification of unimodal methods for depression detection, illustrating the diverse range of techniques employed across different data modalities. **??** illustrates the hierarchical classification of these unimodal methods for depression detection, categorizing approaches into text-based, audio-based, video and facial expression analysis, and social media interactions. Each category is further divided into subcategories detailing specific techniques and advancements, thereby highlighting the diverse methodologies employed in mental health assessments. This visual representation not only enhances our understanding of the classification but also emphasizes the complexity and breadth of unimodal approaches in the context of mental health research.

3.1 Text-Based Approaches

Text-based methods are fundamental in detecting depression, employing linguistic analysis to identify patterns linked to mental health disorders. Advances in natural language processing (NLP) and machine learning have facilitated the examination of textual data from clinical interviews, social

Method Name	Linguistic Analysis	Multimodal Approaches	Advanced Models
MT-BGRU[55]	Text Embeddings	Multimodal Data Integration	Multi-task Bgru
HAN+L[53]	Affective Linguistic Features	-	Hierarchical Attention Network
SSL-DD[56]	Speech Data Patterns	-	Self-supervised Learning
VBGMF[54]	-	Body Gestures	Linear Regression Classifier

Table 2: This table presents a comparative overview of various text-based methods employed in depression detection, highlighting their linguistic analysis, multimodal approaches, and advanced modeling techniques. The table includes models such as MT-BGRU, HAN+L, SSL-DD, and VBGMF, showcasing their unique methodological contributions and integration of NLP, machine learning, and multimodal data.

media, and blogs, revealing correlations between linguistic features and symptoms of depression, anxiety, and suicidal ideation [8]. The MT-BGRU model, leveraging a bidirectional gated recurrent unit architecture with a multi-task loss function, exemplifies recent advancements in predicting depression presence and severity from text by addressing data sparsity in clinical settings [55].

Integrating hierarchical representation learning with external affective information enhances detection accuracy, providing a comprehensive understanding of emotional expressions in text [53]. The EmoMent dataset offers insights into emotional expressions related to mental health [13]. Additionally, self-supervised learning (SSL)-based speech models have emerged to identify individual depression symptoms and predict severity [56]. Combining vision-based body gesture analysis with text data further illustrates the potential of multimodal approaches in predicting non-clinical depression [54].

As illustrated in Figure 2, the hierarchical structure of text-based approaches in depression detection highlights key categories such as linguistic analysis, emotional expressions, and advanced models. Each category encompasses specific methodologies and datasets that contribute to the field, showcasing the integration of NLP, machine learning, and multimodal approaches. Text-based methods, enhanced by NLP and machine learning, significantly improve diagnostic accuracy and scalability. Large language models (LLMs) outperform traditional methods in detecting mental disorders, especially in challenging contexts, while integrating psycholinguistic features with encoder-based models provides valuable resources for practitioners [50, 24, 61]. Table 2 provides a comprehensive comparison of text-based methods used in depression detection, emphasizing their respective approaches in linguistic analysis, multimodal integration, and advanced modeling.

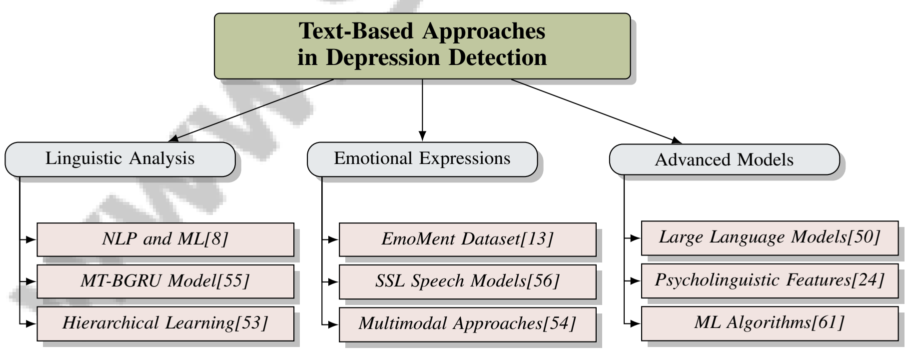


Figure 2: This figure illustrates the hierarchical structure of text-based approaches in depression detection, highlighting key categories such as linguistic analysis, emotional expressions, and advanced models. Each category encompasses specific methodologies and datasets that contribute to the field, showcasing the integration of NLP, machine learning, and multimodal approaches.

3.2 Audio-Based Approaches

Audio-based methods analyze speech acoustic features to detect depression, examining tonal variations, prosodic features, and spectral properties. Deep convolutional neural networks (CNNs) have proven effective in detecting depression from voice samples [16]. The Audio-based Depression Risk Assessment (ADRA) method utilizes telehealth audio data for depression risk prediction, demonstrating the utility of remote assessments [9].

Method Name	Technological Methods	Feature Analysis	Data Integration
ASSD[16]	Deep Convolutional Networks	Vocal Features Analysis	-
ADRA[9]	Machine Learning Model	Audio Features	Multimodal Data
ABAFnet[3]	Lstm-Attention Pipeline	Multiple Acoustic Features	Speech Feature Fusion

Table 3: Overview of audio-based methods for depression detection, highlighting the technological methods, feature analysis, and data integration strategies employed. The table compares three approaches: ASSD, ADRA, and ABAFnet, focusing on their unique methodologies and integration capabilities.

Table 3 provides a comprehensive comparison of various audio-based approaches for detecting depression, detailing the technological methods, feature analysis, and data integration techniques utilized by each method. Attention-based acoustic feature fusion techniques enhance model interpretability and accuracy, addressing subjective and clinical assessment challenges [3]. Deep learning architectures like recurrent neural networks (RNNs) and CNNs facilitate intricate pattern recognition in audio signals, essential for predicting mental disorders. The integration of audio and textual data through embedding techniques, such as WaveNet for audio encoding and transformers like BERT for text, allows comprehensive psychological state analysis [62, 63, 35]. These advancements highlight the potential of audio-based methodologies as robust tools for mental health assessment.

3.3 Video and Facial Expression Analysis

Video-based methods capture non-verbal cues, particularly facial expressions, to assess depression. These methods emphasize facial image analysis during interviews to gauge depression levels [3]. The Multi-stream CNN-GCN Framework (MS-CNN-GCN) enhances depression recognition by processing facial attribute time-series data and optimizing feature extraction through Neural Architecture Search (NAS) techniques [26, 64]. Temporal features, extracted using a Temporal Dilated Convolutional Network (TDCN) with a Feature-Wise Attention (FWA) module, significantly enhance depression detection [1].

Research on remote photoplethysmographic signals (rPPG) from facial videos enables the identification of physiological features associated with depression. This contactless approach assesses mental health by analyzing subtle changes in blood volume pulse (BVP) and heart rate variability (HRV) captured through facial imagery [65, 66, 67, 68, 69]. Video-based methods continue to play a critical role in comprehensive mental health assessment, offering promising avenues for early detection and intervention.

3.4 Social Media and Online Interactions

Method Name	Data Source	Analytical Techniques	Model Objectives
DeSK[60]	Reddit, Weibo	Multi-task Learning	Improved Detection Accuracy
DN[59]	User Tweets	Deep Learning	Improved Accuracy
MDD[58]	Smartphone Usage Data	Feature Extraction	Early Diagnosis
CD[57]	Social Media Content	Llms	Explainability

Table 4: Summary of various methodologies for depression detection utilizing social media and smartphone data. The table outlines the data sources, analytical techniques, and model objectives of different studies, highlighting advancements in detection accuracy, early diagnosis, and explainability.

Social media data provides a promising avenue for depression detection, leveraging user-generated content on platforms like Twitter and Facebook. These platforms offer rich datasets for identifying depressive behaviors through user posts and interactions [60]. The COVID-19 pandemic has intensified expressions of feelings on social media, underscoring their relevance in understanding mental health trends [70]. Table 4 provides a comprehensive overview of the methodologies employed in recent studies for detecting depression through social media and smartphone data, emphasizing the diversity of data sources and analytical approaches.

Models like DepressionNet utilize hybrid summarization strategies to filter user tweets, followed by deep learning models for refined content analysis [59]. Advanced attention mechanisms enhance feature extraction and prioritization from speech data, facilitating effective classification of depressive states [70]. Multimodal approaches, such as the Multimodal Object-Oriented Graph Attention Model

(MOGAM), integrate features from vlogs, demonstrating the benefits of combining multiple data types in mental health assessments [58]. The D-Vlog dataset exemplifies efforts to systematically extract and annotate relevant data for model training and testing, improving depression detection accuracy [71].

Developing explainable models for early depression detection through social media analysis is vital for enhancing interpretability and transparency [57]. Integrating figurative language detection improves the identification of depressive symptoms from tweets, highlighting the complexity of accurately detecting mental health conditions in online interactions [70]. These advancements underscore the potential of leveraging social media data and mobile technologies for mental health assessment, providing valuable tools for early detection and intervention in depression.

Feature	Text-Based Approaches	Audio-Based Approaches	Video and Facial Expression Analysis
Data Type	Textual Data	Audio Data	Video Data
Key Technique	Linguistic Analysis	Acoustic Feature Analysis	Facial Expression Analysis
Unique Feature	Mt-BGRU Model	Attention-based Fusion	Ms-CNN-GCN Framework

Table 5: This table provides a comparative analysis of unimodal methods for depression detection, illustrating the distinct data types and key techniques employed across text, audio, and video modalities. It highlights unique features of each approach, such as the Mt-BGRU model for text, attention-based fusion for audio, and the Ms-CNN-GCN framework for video and facial expression analysis. This comparison underscores the diverse methodologies utilized in mental health assessments.

4 Multimodal Methods for Depression Detection

The integration of diverse data types is pivotal in advancing depression detection methodologies. This section underscores how synthesizing multiple modalities—audio, visual, and textual—enhances the accuracy and robustness of mental health assessments. By leveraging the complementary strengths of these data types, researchers are developing comprehensive frameworks for understanding individual mental states. The following subsection details specific techniques and approaches exemplifying this integration for improved depression detection.

4.1 Integration of Multiple Data Types

Integrating multiple data types in depression detection significantly enhances the accuracy and robustness of mental health assessments. Multimodal approaches leverage the strengths of diverse modalities, such as audio, visual, and textual data, to provide a comprehensive understanding of mental states. Techniques like the Knowledge-infused Neural Network (KiNN) demonstrate the potential of integrating domain-specific and commonsense knowledge to improve detection accuracy through diverse information sources [14]. Similarly, DepressionNet utilizes a hierarchical deep learning approach, combining user behavior features with post summarization to enhance depression detection in social media users [59].

Advanced methods employ state-of-the-art machine learning algorithms to process and fuse data from different modalities. For instance, SSL-DD utilizes self-supervised learning (SSL)-based speech models to extract embeddings from speech data, aiding in symptom identification and severity prediction [56]. The integration of linguistic data into transformer models, as proposed by Ilias et al., further exemplifies enhancing detection accuracy by incorporating diverse data types [71].

Combining speech features with cognitive task performance metrics, like the n-Back Task, showcases the potential of integrating cognitive and acoustic data for improved depression symptom classification [72]. Additionally, modeling short-term facial behaviors using rank pooling and clip-level behaviors with Gaussian Mixture Models highlights the effectiveness of integrating non-verbal cues for depression classification [12].

In social media analysis, methods like DeSK leverage sentiment features, enhancing depression detection by integrating shared sentiment knowledge [60]. The use of deep learning models, such as recurrent neural networks, to analyze textual data without extensive feature engineering underscores the potential of integrating diverse data types for enhanced mental health assessments [73].

AI-generated, for reference only.

The MMFF method exemplifies integrating multi-order factors from different modalities, enhancing depression diagnosis performance by extracting and combining features from diverse data sources [2]. Additionally, ABAFnet combines various acoustic features using a late fusion strategy and an attention mechanism, enhancing detection accuracy [3].

The use of rPPG methods to derive physiological insights from visual information highlights the effectiveness of multimodal approaches in correlating depressive states with physiological signals [66]. The DeepMood architecture, utilizing Gated Recurrent Units (GRU) for feature extraction, models multi-view time series data from keypresses and accelerometer data, showcasing the integration of diverse data types for mental health assessments [32].

The one-shot learning-based Siamese network method models the similarity between audio-textual speech encodings to detect depression relapse, illustrating the potential of integrating audio and textual data for improved detection accuracy [37]. Additionally, segmenting interviews by topics and extracting audio, video, and semantic features exemplifies effective integration of diverse data types [22].

These advancements highlight the dynamic nature of multimodal approaches, paving the way for more effective mental health assessments. By harnessing the synergy between different data types, these techniques significantly improve detection accuracy and reliability, advancing the field of depression detection [5].

4.2 Multimodal Integration and Fusion Techniques

Multimodal integration and fusion techniques are crucial in enhancing the accuracy and reliability of depression detection systems by effectively combining data from various modalities. These methods leverage the strengths of different data types, such as audio, text, and visual cues, to provide a comprehensive analysis of mental health states. The innovative Beam Search Fusion (BS-Fusion) method exemplifies this approach by effectively combining the outputs of multiple classifiers and modalities, enhancing classification performance [74].

The integration process often involves sophisticated fusion strategies operating at different data representation levels. For instance, Hu et al. propose modality fusion at syntactic and semantic levels, combined with inter-modal contrastive learning, to improve sentiment and emotion understanding [75]. This approach underscores the importance of capturing emotional expressions' nuances across modalities.

Attention-based fusion networks have been instrumental in improving depression detection systems' overall estimation accuracy. These networks automatically weight the contributions of different modalities, as demonstrated by Qureshi et al., enhancing multimodal assessments' precision [76]. Similarly, the DepMamba framework employs a progressive fusion strategy that enhances both intermodal and intramodal feature extraction, facilitating a more nuanced understanding of depression-related cues [77].

Early fusion techniques, where features from different modalities are combined at the initial stages of data processing, have shown significant performance improvements compared to unimodal approaches. Deng et al. highlight this approach's effectiveness in their multimodal neural architecture, integrating features from different modalities at the utterance level [78].

Advanced deep learning frameworks enhance the fusion process by incorporating features like Mel-frequency cepstral coefficients (MFCC) with data augmentation and transfer learning techniques, significantly improving depression detection accuracy, as demonstrated by Rejaibi et al. [79]. Moreover, structured approaches like CubeMLP, which mix multimodal features across sequence, modality, and channel axes using independent MLP units, offer a systematic method for feature integration [80].

Analyzing dyadic conversations by embedding multimodal features into transcripts for depression prediction represents another innovative approach to fusion techniques [81]. This method underscores the potential of integrating conversational data with other modalities to enhance mental health assessments.

Finally, the MMFF framework exemplifies extracting and integrating multi-order factors from various modalities through a shared latent proxy, facilitating accurate final predictions [2]. These advancements in multimodal integration and fusion techniques underscore the ongoing evolution of methodologies in depression detection, highlighting the potential for more effective mental health assessments.

As illustrated in Figure 3, the hierarchical structure of multimodal integration techniques highlights key fusion strategies, feature extraction methods, and application areas in depression detection and emotion recognition. This figure presents a comprehensive overview of the various multimodal methods employed in depression detection, emphasizing the integration of diverse data types such as visual, audio, and textual information. By capturing the complex interplay of emotional cues through these techniques, researchers can enhance the detection and understanding of depression, ultimately leading to more effective mental health assessments [74, 75].

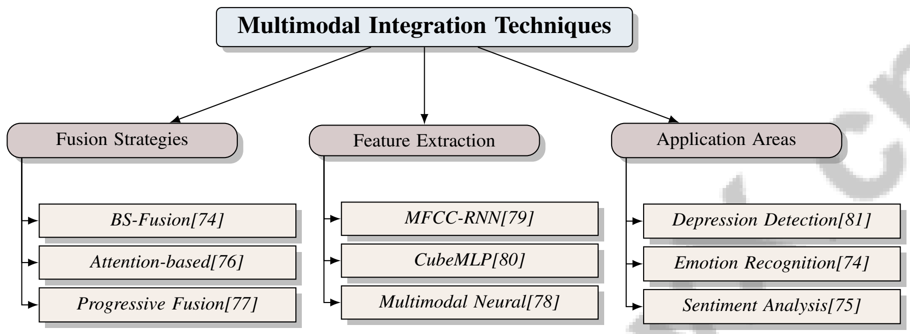


Figure 3: This figure illustrates the hierarchical structure of multimodal integration techniques, highlighting key fusion strategies, feature extraction methods, and application areas in depression detection and emotion recognition.

4.3 Case Studies and Applications

The application of multimodal methodologies in mental health assessment has led to significant advancements in accurately detecting depression and related disorders. A notable case study involves using the Distress Analysis Interview Corpus-Wizard of Oz (DAIC-WOZ) dataset, part of the larger Distress Analysis Interview Corpus and the Remote Collaborative and Affective Interactions database. This dataset comprises real-world data from clinical interviews and collaborative tasks, integrating verbal and non-verbal cues to enhance depression detection accuracy [82]. The integration of these diverse data types underscores the effectiveness of multimodal systems in mental health assessments.

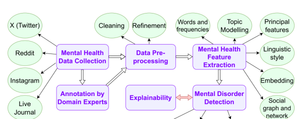
Another study utilizing the Thymia Research platform involved 969 participants, divided into patient and control groups, to gather data on speech and n-Back performance. This study exemplifies the potential of combining cognitive and acoustic data to improve the classification of depression symptoms [72]. The integration of speech features with performance metrics from cognitive tasks highlights the advantages of multimodal approaches in providing a comprehensive understanding of mental health states.

The development of a multimodal detection system for diagnosing ADHD and depression further demonstrates the state-of-the-art performance in integrating audiovisual information. This system exemplifies a successful case study in applying multimodal methodologies for clinical diagnosis, highlighting their potential to enhance predictive performance and improve decision-making in healthcare by integrating diverse data sources, such as acoustic, linguistic, and visual information, particularly in assessing complex conditions like bipolar disorder [17, 83, 84]. The LMVD dataset, representing the largest collection of audiovisual data for depression recognition in everyday contexts, supports the effectiveness of integrating diverse data types in mental health assessments.

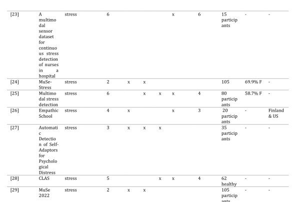
Future research directions suggest exploring the integration of additional modalities, such as images and memes, to enhance the understanding of figurative language in mental health contexts. This approach has the potential to yield comprehensive insights into the complex and varied expressions of mental health conditions by integrating multimodal data sources, such as visual, textual, and audio cues, which reflect the emotional and psychological states of individuals. By analyzing diverse features from social media interactions, including user demographics and language patterns, this method enhances the accuracy of mental health assessments and supports developing targeted interventions that account for these nuances [85, 35, 8, 13, 26]. Additionally, developing methodologies that integrate kinemes with other behavioral markers for depression detection represents a promising avenue for future studies.

The BDAE model, which outperforms existing methods in emotion recognition tasks, exemplifies a viable approach for integrating multimodal data, achieving high accuracy rates in detecting emotional states. The MDD-F framework effectively integrates visual, textual, and connectivity features to enhance the identification of depressive behaviors among Twitter users, as demonstrated by experimental evaluations that improved the average F1-Score by 5 percent over existing state-of-the-art methods. This approach leverages multimodal data and facilitates demographic inference, contributing to developing more targeted mental health interventions [26, 86, 41].

Comparative analyses reveal that multimodal approaches, particularly those combining physiological data with action units, significantly enhance emotion recognition accuracy compared to unimodal methods. This finding underscores the importance of leveraging multiple data sources to improve the reliability and depth of mental health assessments. The case studies and applications presented illustrate the significant impact of multimodal methodologies—integrating visual, textual, and interaction data—on enhancing the accuracy and effectiveness of depression detection and mental health assessment. By leveraging diverse data sources from social media and employing advanced statistical techniques, these approaches improve the identification of depressive behaviors and facilitate demographic insights that can inform targeted health interventions. This comprehensive framework addresses the limitations of traditional unimodal analysis and underscores the importance of a holistic view in understanding and diagnosing mental health conditions [26, 41, 87, 88].

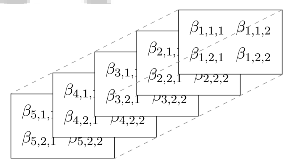


(a) Mental Health Data Collection and Analysis Framework[23]



[23]	A multimodal sensor dataset for continuous stress detection of nurses in a hospital	stress	6				x	6	15 participants	-	-
[24]	MuSe-Stress	stress	2	x	x				105	69.9% F	-
[25]	Multimodal stress detection	stress	6		x	x	x	4	80 participants	58.7% F	-
[26]	Empathic School	stress	4	x			x	3	20 participants	-	Finland & US
[27]	Automatic Detection of Self-Adaptors for Psychological Distress	stress	3	x	x	x			35 participants	-	-
[28]	CLAS	stress	5			x	x	4	62 healthy	-	-
[29]	MuSe 2022	stress	2	x	x				105 participants	-	-

(b) A Multimodal Sensor Dataset for Continuous Stress Detection of Nurses in a Hospital[84]



(c) The image depicts a hierarchical structure of boxes and rectangles, with dashed lines connecting them, forming a tree-like pattern.[89]

Figure 4: Examples of Case Studies and Applications

As shown in Figure 4, multimodal methods in mental health research, particularly in detecting and analyzing depression, are gaining traction as a promising approach. The figure showcases three distinct examples of case studies and applications highlighting these methods' versatility and effectiveness. The first example, "Mental Health Data Collection and Analysis Framework," illustrates a comprehensive process integrating data from various social media platforms, emphasizing data preprocessing and feature extraction in accurately detecting mental disorders. The second example, "A Multimodal Sensor Dataset for Continuous Stress Detection of Nurses in a Hospital," underscores the practical application of multimodal datasets in real-world settings, such as hospitals, where continuous monitoring of stress levels in nurses is vital for maintaining their well-being and ensuring high-quality patient care. Lastly, the hierarchical structure depicted in the third example provides a visual representation of categorization and classification processes, crucial in organizing and interpreting complex data sets. Together, these examples demonstrate the potential of multimodal methods to enhance our understanding and detection of depression, offering innovative solutions for mental health challenges [23, 84, 89].

5 Emotion Recognition and Affective Computing

The accurate identification and interpretation of emotional states are crucial in emotion recognition and affective computing, particularly for mental health assessments. This section explores foundational aspects of emotion recognition, focusing on facial and vocal modalities. Understanding these modalities' functions and significance enhances mental health evaluations and interventions.

5.1 Facial and Vocal Emotion Recognition

Facial and vocal emotion recognition are vital in assessing mental health, providing insights into emotional states indicative of conditions like depression. Capturing and interpreting subtle cues from facial expressions and vocal tones is essential for accurate assessments. The Chat-Diagnose system exemplifies interactive dialogues to enhance understanding of mental states, highlighting interaction's role in emotion recognition [57]. The EmoViz dashboard demonstrates neural networks' application in analyzing speech signals for emotion classification, underscoring speech-based emotion recognition systems' potential [33]. Multimodal approaches, such as those by Franceschini et al., align emotional sequence representations across modalities using contrastive loss, emphasizing facial and vocal cues' integration for comprehensive recognition [90]. Datasets like Priori, annotating audio for emotional content, support training robust models for recognizing speech-based emotions [91]. Advances in AI algorithms, as noted by Bhatt et al., enhance cognitive behavior detection by integrating diverse data sources, refining emotion analysis [19]. Ongoing advancements in multimodal deep learning approaches, analyzing audio and text features, improve mental health assessments' precision and efficacy. These innovations leverage embedding techniques and emotion-annotated corpora to detect disorders like depression and schizophrenia by integrating linguistic and emotional cues [36, 13, 63, 35]. By harnessing different modalities' strengths, these methods offer tools for early detection and intervention in mental health disorders.

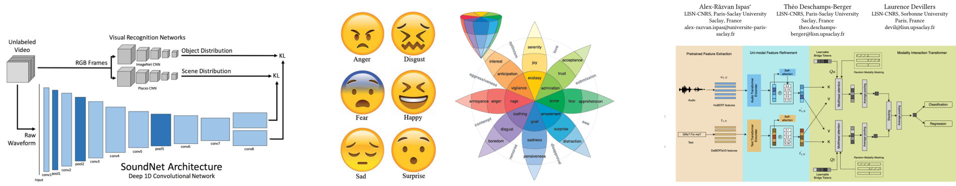
5.2 Role of Linguistic and Acoustic Features

Linguistic and acoustic features are pivotal in emotion recognition, offering insights into emotional states underlying disorders like depression. Integrating these features into models allows nuanced emotional expression analysis, enhancing assessments' accuracy. Analyzing linguistic and contextual features from tweets offers a promising avenue for detecting early depression and PTSD signs, demonstrating linguistic analysis's potential in mental health monitoring [92]. Current reliance on lexicon-based tools presents limitations in capturing context and subtleties, leading to sentiment misinterpretation, highlighting the need for sophisticated models [93]. Integrating linguistic features into advanced models offers a comprehensive understanding of emotional expressions. Acoustic features from vocal data are equally crucial. EmoViz exemplifies machine learning models analyzing speech for emotion identification, showcasing acoustic data's potential in assessments [33]. Levinson et al. emphasize capturing vocal nuances correlating with depression, highlighting acoustic features' role in identifying emotional states [9]. Integrating linguistic and acoustic features into systems improves mental health assessments' accuracy by leveraging natural language processing and multimodal deep learning. These approaches analyze spoken language content and emotional cues in voice recordings, providing a nuanced understanding of conditions. Combining text analysis with audio processing enhances disorder detection, demonstrated by studies using emotion-annotated corpora and embedding models. This framework facilitates early detection and intervention, addressing human expression complexities [13, 35, 36]. By leveraging modalities' strengths, these methods offer tools for early detection and intervention in mental health disorders.

5.3 Applications of Emotion Recognition in Real-World Scenarios

Emotion recognition technologies have practical applications in real-world settings, enhancing the assessment and understanding of disorders like depression. These technologies use advanced models to analyze emotional expressions from various sources, providing insights for professionals and researchers. The EmoViz tool offers a dashboard visualizing emotion analyses, clustering audio and text based on emotions [33]. This capability allows continuous emotional state monitoring, facilitating timely interventions and improved outcomes. LSTM-Att model's performance in identifying emotions in Tweets demonstrates emotion recognition technologies' potential in social media analysis [94]. By detecting emotional content in interactions, these models provide insights into users' mental health, enabling early symptom detection and personalized interventions. Morini et al. highlight emotional trajectories' complexity in online interactions, revealing non-linear progressions in users discussing depression [95]. This challenges traditional engagement models, suggesting emotion recognition technologies' crucial role in understanding mental health experiences' dynamic nature. Practical applications emphasize technologies' potential to transform assessment and intervention. By utilizing insights from AI and machine learning, professionals gain a deeper understanding of emotional needs. This aids in tailoring effective interventions, such as internet-delivered Cognitive Behavioral

Therapy, fostering improved engagement through trained coaches, leading to better outcomes and care quality. Integrating explainable AI models in analytics allows transparency and interpretability in decision-making, addressing ethical considerations and guiding future research and practice [23, 6].



(a) Deep Learning Framework for Visual and Sound Recognition in Unlabeled Videos[74]

(b) Emotions and Emojis[40]

(c) Audio-Text Fusion for Multimodal Speech Recognition[96]

Figure 5: Examples of Applications of Emotion Recognition in Real-World Scenarios

As illustrated in Figure 5, emotion recognition and affective computing enhance human-computer interaction by enabling machines to understand and respond to emotions. Real-world applications are showcased through various frameworks and systems. A deep learning framework integrates visual and sound recognition to interpret objects and scenes in videos, leveraging networks trained on datasets like ImageNet and Places. This approach underscores combining visual and auditory data to enrich recognition. Emojis represent emotions, offering a simplified yet effective way to convey complex states, bridging digital communication and human feelings. Audio and text fusion in multimodal speech recognition systems demonstrates advancements in refining cues from different sources, enhancing detection accuracy and context. Collectively, these examples highlight diverse methodologies and applications in real-world scenarios, illustrating growing importance in creating intuitive and responsive environments [74, 40, 96].

5.4 Advancements in Emotion Recognition Technologies

Recent advancements in emotion recognition technologies have improved systems' accuracy and applicability across domains. Deep learning algorithms enhance emotion recognition across modalities, evidenced by significant progress in current research [20]. These advancements are facilitated by sophisticated models capturing complex expressions, improving systems' accuracy. Meta features, as demonstrated by Orton et al., enhance psychological distress prediction, suggesting advancements' crucial role in refining technologies [54]. Accurately predicting emotional states from diverse sources underscores integrating cutting-edge technologies' transformative potential. The DeepMood architecture exemplifies mobile phone metadata's feasibility in inferring mood disturbances, achieving 90.31

6 Deep Learning and Neural Networks in Mental Health Assessment

The integration of deep learning into mental health assessment is rapidly advancing, highlighting its transformative impact on depression detection. This section examines the applications of these models, emphasizing their role in enhancing mental health evaluations and interventions.

6.1 Applications of Deep Learning Models

Deep learning models have significantly advanced depression detection by analyzing complex data across multiple modalities, such as audio, text, and visual inputs, enhancing assessment accuracy and interpretability. Integrating diverse data representations with energy consumption analysis effectively balances classification accuracy and computational efficiency [97]. Large language models (LLMs) outperform traditional methods, especially on noisy datasets, underscoring their potential in mental health assessments [50]. The Chat-Diagnose method exemplifies LLMs' effectiveness in leveraging semantic understanding for depression detection [57].

In speech emotion recognition, deep learning surpasses existing methods, enhancing intelligent virtual assistants (IVAs) and supporting mental health monitoring [34]. Chain-of-Thought (CoT) prompting improves AI model accuracy in predicting PHQ-8 scores, showcasing the potential of innovative

AI-generated, for reference only.

prompting techniques [29]. Augmentation methods enhance depression detection accuracy, with significant improvements in classification performance [98]. The Automatic Depression Detection (ADD) method demonstrates state-of-the-art performance using Temporal Dilated Convolutional Networks (TDCN) and Feature-Wise Attention (FWA) for detecting depression from visual cues [15].

The SEO framework improves task completion and emotional support in depression diagnosis, emphasizing empathetic frameworks' role in assessments [99]. Evaluating viewers' mental health through classifiers trained with Naive Bayes and LSTM models, validated by CES-D scores from video comments, highlights deep learning's diverse applications [100]. Integrating external affective information into hierarchical networks achieves state-of-the-art results in depression detection accuracy [53]. The Behavioral Depression Degree (BDD) correlates strongly with traditional scales, effectively assessing depression severity through multimodal methods [7].

The deployment of deep learning models represents a significant advancement in mental health assessment. By harnessing these models, researchers and clinicians achieve more accurate, interpretable, and generalizable solutions for identifying depressive symptoms across diverse populations and modalities. Recent advancements, particularly through CoT prompting and multimodal data analysis, revolutionize mental health interventions. These technologies enhance depression assessment accuracy, enabling a nuanced understanding of individual mental health conditions. Integrating visual, textual, and connectivity clues from social media allows comprehensive analyses of depressive behaviors and demographic factors, improving risk identification. Collectively, these innovations refine diagnostic processes and support developing targeted, demographic-aware mental health interventions, leading to better outcomes for individuals experiencing depression and related disorders [26, 29].

6.2 Innovative Techniques and Frameworks

Recent advancements in deep learning have introduced innovative frameworks and techniques significantly enhancing mental health assessment, particularly in depression detection. Rodríguez et al. introduced a two-stage approach predicting individual depression symptoms using CNN-LSTM models, emphasizing understanding granular aspects for a nuanced assessment [101]. Xu et al. proposed a two-stage temporal modeling framework incorporating a Depression Feature Enhancement (DFE) module and novel graph encoding strategies, capturing complex temporal dynamics of depression symptoms [67].

In social media analysis, the TEDD approach employs transformer-based classifiers fine-tuned on depression-related data, effectively predicting depression levels [102]. The SACB framework uses a Convolutional Bidirectional LSTM architecture for end-to-end depression estimation, optimizing feature extraction from diverse data sources [49]. Gu et al. propose simultaneous training of feature extraction and fusion modules, enhancing model performance through optimal global parameter tuning [103].

Transformer-based approaches with late-fusion strategies integrate ECG and EEG signals, exemplifying innovative use in multimodal emotion recognition [104]. These advancements highlight the versatility of transformer models in handling complex physiological data, paving the way for comprehensive mental health assessments.

Recent studies underscore that advancements in deep learning methodologies are transforming mental health assessment practices. Techniques such as LLMs, hybrid summarization strategies, and multimodal approaches demonstrate enhanced accuracy in detecting depression and psychological disorders. LLMs perform well on noisy datasets, while frameworks like DepressionNet improve detection rates on social media platforms. Multimodal approaches integrating facial recognition and dynamic textual analysis effectively distinguish varying depression levels and related conditions. These developments offer promising avenues for early diagnosis and targeted interventions, enhancing mental health care [50, 59, 30, 105]. By harnessing advanced architectures and integrated training strategies, researchers continue to push the boundaries in mental health technology.

6.3 Challenges in Deep Learning for Mental Health

Deep learning models in mental health applications face several challenges impacting their effectiveness and real-world applicability. A primary concern is the lack of explainability and interpretability, crucial for clinical adoption where transparency and understanding of AI decisions are paramount [106]. The intricate nature of deep neural networks often results in "black box" models, complicating clinicians' ability to trust their outputs and hindering AI integration [23].

Another significant challenge arises from dependence on web-scraped and social media data, which may introduce biases and inaccuracies due to variability and inconsistency in online behavior [4]. The diversity in language use across platforms complicates accurate interpretation, as these data sources may not reliably reflect an individual's mental health status [107]. This variability, combined with noise from segment-level labeling, complicates the practical deployment of models [106].

Generalization remains a critical issue, with models often exhibiting inconsistent performance across datasets and populations. This problem is exacerbated by the limited availability of large-scale, diverse datasets necessary for effective generalization [108]. The homogeneous nature of many existing datasets may result in models that do not represent broader populations, limiting applicability [109].

The integration of multimodal data adds complexity, requiring models to effectively combine diverse data types such as text, audio, and visual cues to provide a holistic understanding [108]. The absence of standardized methods for integrating these modalities can impede developing robust models capable of capturing the intricacies of human emotions and psychological conditions [2].

Handling linguistic nuances and complexities of human language poses additional challenges in classification tasks. Models that oversimplify language patterns risk missing critical indicators, leading to inaccurate assessments [110]. Addressing these challenges necessitates ongoing research and development to improve interpretability, generalizability, and integration capabilities, ultimately enhancing effectiveness in clinical settings.

7 Challenges and Future Directions

Addressing the limitations of current depression detection methodologies is essential for advancing their efficacy. This section examines the challenges impeding these technologies and explores future research directions to enhance their effectiveness.

7.1 Limitations and Challenges

The adoption of deep learning models in clinical settings is hindered by their opaque nature, necessitating frameworks that clarify AI decision-making processes [14]. The scarcity of comprehensive datasets limits model generalizability across diverse populations, exacerbated by reliance on self-reported measures and the absence of objective biomarkers [31, 72]. Data imbalance, particularly in social media analysis, further restricts applicability [18, 73]. The complexity of human emotions and lack of multimodal datasets compound these challenges, complicating data acquisition and analysis [19, 2].

Technological constraints, such as the need for high-quality video in rPPG methods and the variability of smartphone usage data, limit the reliability of certain approaches [66, 32]. Establishing standardized protocols and regulatory frameworks, grounded in tools like the PHQ-9, is crucial for improving the reliability and generalizability of these technologies [111, 112].

7.2 Future Directions and Research Opportunities

Future research should focus on developing multimodal systems that integrate diverse data types to enhance model accuracy [55]. Exploring transfer learning and domain adaptation can improve model generalization across populations [71]. Self-supervised learning (SSL) models also hold promise for detecting less correlated symptoms [56]. Enhancing rPPG methods and integrating additional modalities can improve accuracy in diverse settings [66]. Larger-scale studies are needed to improve dataset quality and explore ethical implications, ensuring responsible use of emotion recognition technologies [20, 6].

Optimizing models for clinical settings and exploring early detection applications are critical areas for future research [3]. Addressing these opportunities can significantly advance depression detection, leading to more accurate and culturally sensitive assessments.

7.3 Data Quality and Availability

The scarcity of large, diverse datasets limits model generalization, exacerbated by biases in self-reported measures and inconsistencies in data sources [73, 72]. Data imbalance, particularly in social media analysis, skews model performance [18]. The integration of multimodal data presents an opportunity to enhance dataset quality but poses challenges in data acquisition and fusion [19, 2].

Developing large-scale, high-quality datasets with standardized protocols is essential for enhancing the generalizability and effectiveness of depression detection technologies. Systematic categorization of online expressions of psychological distress, as demonstrated by the EmoMent corpus, is critical for understanding the relationship between demographics and mental health [26, 13].

7.4 Model Generalization and Interpretability

Model generalization is hindered by insufficient labeled data and reliance on irrelevant features, reducing accuracy in detecting depressive symptoms [30]. The integration of multimodal data necessitates sophisticated methods for data fusion, yet standardized protocols are lacking [113]. Interpretability remains a critical issue, as "black box" models lack transparency, impeding clinical adoption [17, 27].

Future research should prioritize developing techniques that enhance generalization and transparency, such as integrating clinically relevant frameworks and exploring explainable AI methodologies [111, 50]. This includes exploring transfer learning and domain adaptation to improve model robustness and utilizing explainable AI frameworks to provide interpretable insights into model decisions.

7.5 Privacy and Ethical Considerations

Ensuring privacy and ethical considerations in mental health assessment is essential for responsible use of sensitive data. Federated learning models, like FedMood, enhance prediction accuracy while preserving privacy [38]. Transparency and ethical data treatment are fundamental, as demonstrated in studies on empathetic depression diagnosis-oriented chats [99]. Utilizing social media data introduces unique ethical challenges, necessitating adherence to privacy guidelines and consideration of psychological impacts [13].

Addressing privacy and ethical considerations ensures responsible implementation of advanced technologies, fostering trust among stakeholders and promoting effective mental health interventions [114, 23].

7.6 Real-World Application and Deployment

Deploying depression detection technologies in real-world settings requires robust, explainable models to gain clinicians' and patients' trust [84]. Ethical deployment necessitates data privacy and user consent, with federated learning models offering a promising solution [84]. Addressing potential biases in model performance and ensuring infrastructure availability are crucial for integrating these technologies into healthcare systems [17, 23].

By addressing these challenges, researchers can enhance the deployment of advanced technologies in real-world settings, improving early detection and understanding of depression and related disorders [111, 30].

8 Conclusion

The survey highlights the pivotal influence of advanced technologies in revolutionizing mental health assessment, with a focus on depression detection and management. By employing multimodal methodologies that amalgamate audio, visual, and textual data, the accuracy and dependability of depression detection systems are markedly improved. Approaches such as the TOAT model adeptly tackle issues of data scarcity and organization, enhancing performance and underscoring the importance of harnessing the diverse strengths of multiple modalities in crafting comprehensive mental health evaluation tools.

The effectiveness of digital human-computer interaction platforms within clinical environments is noteworthy, with these tools proving as effective as traditional psychiatric interviews for depression screening. The InterMind system exemplifies the benefits of integrating family participation with sophisticated language models to boost diagnostic precision and efficiency, highlighting technology's crucial role in advancing mental health assessments.

Neural networks demonstrate significant potential in providing objective depression diagnoses, achieving commendable accuracy through the analysis of physical signals. The deployment of deep learning-based chatbots, such as Evebot, has yielded notable improvements in users' emotional states, showcasing their capacity to foster positive emotions and mitigate stress. Moreover, the SDCNL method underscores the potential for real-world diagnostic applications in mental health by substantially enhancing classification accuracy following label correction.

The survey also emphasizes the necessity of methodological rigor in machine learning research focused on depression detection via social media to ensure dependable results. The integration of sentiment analysis with multi-task learning, as illustrated by the DeSK method, highlights the importance of augmenting depression detection capabilities. Additionally, the superior performance of models like MultiEEG-GPT in predicting mental health conditions attests to the advantages of multimodal data integration.

References

[1] Nadee Seneviratne and Carol Espy-Wilson. Speech based depression severity level classification using a multi-stage dilated cnn-lstm model, 2021.

[2] Chengbo Yuan, Qianhui Xu, and Yong Luo. Depression diagnosis and analysis via multimodal multi-order factor fusion, 2022.

[3] Xiao Xu, Yang Wang, Xinru Wei, Fei Wang, and Xizhe Zhang. Attention-based acoustic feature fusion network for depression detection, 2023.

[4] Ayaan Haque, Viraaj Reddi, and Tyler Giallanza. Deep learning for suicide and depression identification with unsupervised label correction, 2021.

[5] Evgeny Stepanov, Stephane Lathuiliere, Shammur Absar Chowdhury, Arindam Ghosh, Radu-Laurentiu Vieriu, Nicu Sebe, and Giuseppe Riccardi. Depression severity estimation from multiple modalities, 2017.

[6] Anja Thieme. Understanding the information needs and practices of human supporters of an online mental health intervention to inform machine learning applications, 2021.

[7] Dongdong Liu, Bowen Liu, Tao Lin, Guangya Liu, Guoyu Yang, Dezhen Qi, Ye Qiu, Yuer Lu, Qinmei Yuan, Stella C Shuai, et al. Measuring depression severity based on facial expression and body movement using deep convolutional neural network. *Frontiers in psychiatry*, 13:1017064, 2022.

[8] B. ODea, T. W. Boonstra, M. E. Larsen, T. Nguyen, S. Venkatesh, and H. Christensen. The relationship between linguistic expression and symptoms of depression, anxiety, and suicidal thoughts: A longitudinal study of blog content, 2018.

[9] Adam Valen Levinson, Abhay Goyal, Roger Ho Chun Man, Roy Ka-Wei Lee, Koustuv Saha, Nimay Parekh, Frederick L. Altice, Lam Yin Cheung, Munmun De Choudhury, and Navin Kumar. Using audio data to facilitate depression risk assessment in primary health care, 2023.

[10] Zhiyuan Zhou, Jilong Liu, Sanwang Wang, Shijie Hao, Yanrong Guo, and Richang Hong. Intermind: A doctor-patient-family interactive depression assessment system empowered by large language models, 2024.

[11] Junjie Yin, Zixun Chen, Kelai Zhou, and Chongyuan Yu. A deep learning based chatbot for campus psychological therapy, 2019.

[12] Chuang Yu. Non-verbal facial action units-based automatic depression classification, 2022.

[13] Thushari Atapattu, Mahen Herath, Charitha Elvitigala, Piyanjali de Zoysa, Kasun Gunawardana, Menasha Thilakaratne, Kasun de Zoysa, and Katrina Falkner. Emoment: An emotion annotated mental health corpus from two south asian countries, 2022.

[14] Sumit Dalal, Sarika Jain, and Mayank Dave. Deep knowledge-infusion for explainable depression detection, 2024.

[15] Yanrong Guo, Chenyang Zhu, Shijie Hao, and Richang Hong. Automatic depression detection via learning and fusing features from visual cues, 2022.

[16] Karol Chlasta, Krzysztof Wołk, and Izabela Krejtz. Automated speech-based screening of depression using deep convolutional neural networks, 2019.

[17] Adrienne Kline, Hanyin Wang, Yikuan Li, Saya Dennis, Meghan Hutch, Zhenxing Xu, Fei Wang, Feixiong Cheng, and Yuan Luo. Multimodal machine learning in precision health, 2022.

[18] Alireza Afzal Aghaei and Nadia Khodaei. Automated depression recognition using multimodal machine learning: A study on the daic-woz dataset. *Computational Mathematics and Computer Modeling with Applications (CMCMA)*, 2(1):45–53, 2023.

[19] Priya Bhatt, Amanrose Sethi, Vaibhav Tasgaonkar, Jugal Shroff, Isha Pendharkar, Aditya Desai, Pratyush Sinha, Aditya Deshpande, Gargi Joshi, Anil Rahate, et al. Machine learning for cognitive behavioral analysis: datasets, methods, paradigms, and research directions. *Brain informatics*, 10(1):18, 2023.

[20] Chiara Zucco, Barbara Calabrese, and Mario Cannataro. Emotion mining: From unimodal to multimodal approaches. In *Brain-Inspired Computing: 4th International Workshop, BrainComp 2019, Cetraro, Italy, July 15–19, 2019, Revised Selected Papers 4*, pages 143–158. Springer, 2021.

[21] Marwan Dhuheir, Abdullatif Albaseer, Emna Baccour, Aiman Erbad, Mohamed Abdallah, and Mounir Hamdi. Emotion recognition for healthcare surveillance systems using neural networks: A survey, 2021.

[22] Yuan Gong and Christian Poellabauer. Topic modeling based multi-modal depression detection, 2018.

[23] Yusif Ibrahimov, Tarique Anwar, and Tommy Yuan. Explainable ai for mental disorder detection via social media: A survey and outlook, 2024.

[24] Adil Rajput and Samara Ahmed. Making a case for social media corpus for detecting depression, 2019.

[25] Ana-Maria Bucur, Andreea-Codrina Moldovan, Krutika Parvatikar, Marcos Zampieri, Ashiqur R. KhudaBukhsh, and Liviu P. Dinu. On the state of nlp approaches to modeling depression in social media: A post-covid-19 outlook, 2024.

[26] Amir Hossein Yazdavar, Mohammad Saeid Mahdavinejad, Goonmeet Bajaj, William Romine, Amirhassan Monadjemi, Krishnaprasad Thirunarayan, Amit Sheth, and Jyotishman Pathak. Fusing visual, textual and connectivity clues for studying mental health, 2019.

[27] Junwei Kuang, Jiaheng Xie, and Zhijun Yan. What symptoms and how long? an interpretable ai approach for depression detection in social media, 2023.

[28] Agnieszka Wołk, Karol Chlasta, and Paweł Holas. Hybrid approach to detecting symptoms of depression in social media entries, 2021.

[29] Elysia Shi, Adithri Manda, London Chowdhury, Runeema Arun, Kevin Zhu, and Michael Lam. Enhancing depression diagnosis with chain-of-thought prompting, 2024.

[30] Wenli Zhang, Jiaheng Xie, Zhu Zhang, and Xiang Liu. Depression detection using digital traces on social media: A knowledge-aware deep learning approach, 2023.

[31] Milena Cukic Radenkovic. Machine learning approaches in detecting the depression from resting-state electroencephalogram (eeg): A review study, 2019.

[32] Bokai Cao, Lei Zheng, Chenwei Zhang, Philip S. Yu, Andrea Piscitello, John Zulueta, Olu Ajilore, Kelly Ryan, and Alex D. Leow. Deepmood: Modeling mobile phone typing dynamics for mood detection, 2018.

[33] Jumana Almahmoud and Kruthika Kikkeri. Speech-based emotion recognition using neural networks and information visualization, 2020.

[34] Nelly Elsayed, Zag ElSayed, Navid Asadizanjani, Murat Ozer, Ahmed Abdelgawad, and Magdy Bayoumi. Speech emotion recognition using supervised deep recurrent system for mental health monitoring, 2022.

[35] Habibeh Naderi, Behrouz Haji Soleimani, and Stan Matwin. Multimodal deep learning for mental disorders prediction from audio speech samples. *arXiv preprint arXiv:1909.01067*, 2019.

[36] Nasser Ghadiri, Rasoul Samani, and Fahime Shahrokh. Integration of text and graph-based features for detecting mental health disorders from voice, 2022.

[37] Alice Othmani and Muhammad Muzammel. An ambient intelligence-based approach for longitudinal monitoring of verbal and vocal depression symptoms, 2023.

[38] Yongquan Hu, Shuning Zhang, Ting Dang, Hong Jia, Flora D. Salim, Wen Hu, and Aaron J. Quigley. Exploring large-scale language models to evaluate eeg-based multimodal data for mental health, 2024.

[39] Aditya Parikh, Misha Sadeghi, and Bjorn Eskofier. Exploring facial biomarkers for depression through temporal analysis of action units, 2024.

[40] Yan Wang, Wei Song, Wei Tao, Antonio Liotta, Dawei Yang, Xinlei Li, Shuyong Gao, Yixuan Sun, Weifeng Ge, Wei Zhang, and Wenqiang Zhang. A systematic review on affective computing: Emotion models, databases, and recent advances, 2022.

[41] Ziheng Zhang, Weizhe Lin, Mingyu Liu, and Marwa Mahmoud. Multimodal deep learning framework for mental disorder recognition. In *2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020)*, pages 344–350. IEEE, 2020.

[42] Shubham Dham, Anirudh Sharma, and Abhinav Dhall. Depression scale recognition from audio, visual and text analysis, 2017.

[43] Wei Liu, Wei-Long Zheng, and Bao-Liang Lu. Multimodal emotion recognition using multimodal deep learning, 2016.

[44] Yuanzhe Huang, Saurab Faruque, Minjie Wu, Akiko Mizuno, Eduardo Diniz, Shaolin Yang, George Dewitt Stetten, Noah Schweitzer, Hecheng Jin, Linghai Wang, and Howard J. Aizenstein. Leveraging the finite states of emotion processing to study late-life mental health, 2024.

[45] Guansong Pang, Ngoc Thien Anh Pham, Emma Baker, Rebecca Bentley, and Anton van den Hengel. Deep depression prediction on longitudinal data via joint anomaly ranking and classification, 2022.

[46] Weizhe Lin, Indigo Orton, Qingbiao Li, Gabriela Pavarini, and Marwa Mahmoud. Looking at the body: Automatic analysis of body gestures and self-adaptors in psychological distress, 2020.

[47] Pavani Chowdary, Bhavyajeet Singh, Rajat Agarwal, and Vinoo Alluri. Lyrically speaking: Exploring the link between lyrical emotions, themes and depression risk, 2024.

[48] Nadee Seneviratne and Carol Espy-Wilson. Multimodal depression classification using articulatory coordination features and hierarchical attention based text embeddings, 2022.

[49] Ping-Cheng Wei, Kunyu Peng, Alina Roitberg, Kailun Yang, Jiaming Zhang, and Rainer Stiefelhagen. Multi-modal depression estimation based on sub-attentional fusion, 2022.

[50] Gleb Kuzmin, Petr Strepetov, Maksim Stankevich, Artem Shelmanov, and Ivan Smirnov. Mental disorders detection in the era of large language models, 2024.

[51] Simmi Marina Joseph, Salvatore Citraro, Virginia Morini, Giulio Rossetti, and Massimo Stella. Cognitive network science quantifies feelings expressed in suicide letters and reddit mental health communities, 2021.

[52] Marcel Trotzek, Sven Koitka, and Christoph M. Friedrich. Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences, 2018.

[53] D. Xezonaki, G. Paraskevopoulos, A. Potamianos, and S. Narayanan. Affective conditioning on hierarchical networks applied to depression detection from transcribed clinical interviews, 2020.

[54] Indigo J. D. Orton. Vision based body gesture meta features for affective computing, 2020.

[55] Heinrich Dinkel, Mengyue Wu, and Kai Yu. Text-based depression detection on sparse data, 2020.

[56] Sri Harsha Dumpala, Katerina Dikaios, Abraham Nunes, Frank Rudzicz, Rudolf Uher, and Sageev Oore. Self-supervised embeddings for detecting individual symptoms of depression, 2024.

[57] Wei Qin, Zetong Chen, Lei Wang, Yunshi Lan, Weijieying Ren, and Richang Hong. Read, diagnose and chat: Towards explainable and interactive llms-augmented depression detection in social media, 2023.

[58] Ravi Prasad Thati, Abhishek Singh Dhadwal, Praveen Kumar, and Sainaba P. A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms. *Multimedia Tools and Applications*, 82(4):4787–4820, 2023.

[59] Hamad Zogan, Imran Razzak, Shoaib Jameel, and Guandong Xu. Depressionnet: A novel summarization boosted deep framework for depression detection on social media, 2021.

[60] Yan Shi, Yao Tian, Chengwei Tong, Chunyan Zhu, Qianqian Li, Mengzhu Zhang, Wei Zhao, Yong Liao, and Pengyuan Zhou. Detect depression from social networks with sentiment knowledge sharing, 2023.

[61] Giuliano Lorenzoni, Cristina Tavares, Nathalia Nascimento, Paulo Alencar, and Donald Cowan. Assessing ml classification algorithms and nlp techniques for depression detection: An experimental case study, 2024.

[62] Panagiotis Tzirakis, George Trigeorgis, Mihalis A. Nicolaou, Björn Schuller, and Stefanos Zafeiriou. End-to-end multimodal emotion recognition using deep neural networks, 2017.

[63] Habibeh Naderi, Behrouz Haji Soleimani, and Stan Matwin. Multimodal deep learning for mental disorders prediction from audio speech samples, 2020.

[64] Mingzhe Chen, Xi Xiao, Bin Zhang, Xinyu Liu, and Runiu Lu. Neural architecture searching for facial attributes-based depression recognition, 2022.

[65] Subigya Nepal, Arvind Pillai, Weichen Wang, Tess Griffin, Amanda C. Collins, Michael Heinz, Damien Lekkas, Shayan Mirjafari, Matthew Nemesure, George Price, Nicholas C. Jacobson, and Andrew T. Campbell. Moodcapture: Depression detection using in-the-wild smartphone images, 2024.

[66] Constantino Álvarez Casado, Manuel Lage Cañellas, and Miguel Bordallo López. Depression recognition using remote photoplethysmography from facial videos, 2022.

[67] Jiaqi Xu, Siyang Song, Keerthy Kusumam, Hatice Gunes, and Michel Valstar. Two-stage temporal modelling framework for video-based depression recognition using graph representation, 2021.

[68] Ruiqi Wang, Jinyang Huang, Jie Zhang, Xin Liu, Xiang Zhang, Zhi Liu, Peng Zhao, Sigui Chen, and Xiao Sun. Facialpulse: An efficient rnn-based depression detection via temporal facial landmarks, 2024.

[69] Monika Gahalawat, Raul Fernandez Rojas, Tanaya Guha, Ramanathan Subramanian, and Roland Goecke. Explainable depression detection via head motion patterns, 2023.

[70] Ana-Maria Bucur, Ioana R. Podină, and Liviu P. Dinu. A psychologically informed part-of-speech analysis of depression in social media, 2021.

[71] Loukas Ilias, Spiros Mouzakitis, and Dimitris Askounis. Calibration of transformer-based models for identifying stress and depression in social media, 2023.

[72] Salvatore Fara, Stefano Goria, Emilia Molimpakis, and Nicholas Cummins. Speech and the n-back task as a lens into depression. how combining both may allow us to isolate different core symptoms of depression, 2022.

[73] Sultan Ahmed, Salman Rakin, Mohammad Washeef Ibn Waliur, Nuzhat Binte Islam, Billal Hossain, and Md. Mostofa Akbar. Depression detection from social media bangla text using recurrent neural networks, 2024.

[74] Zheng Lian, Ya Li, Jianhua Tao, and Jian Huang. Investigation of multimodal features, classifiers and fusion methods for emotion recognition, 2018.

[75] Guimin Hu, Ting-En Lin, Yi Zhao, Guangming Lu, Yuchuan Wu, and Yongbin Li. Unimse: Towards unified multimodal sentiment analysis and emotion recognition, 2022.

[76] Syed Arbaaz Qureshi, Sriparna Saha, Mohammed Hasanuzzaman, and Gaël Dias. Multitask representation learning for multimodal estimation of depression level. *IEEE Intelligent Systems*, 34(5):45–52, 2019.

[77] Jiaxin Ye, Junping Zhang, and Hongming Shan. Depmamba: Progressive fusion mamba for multimodal depression detection, 2024.

[78] Didan Deng, Yuqian Zhou, Jimin Pi, and Bertram E. Shi. Multimodal utterance-level affect analysis using visual, audio and text features, 2018.

[79] Emna Rejaibi, Ali Komaty, Fabrice Meriaudeau, Said Agrebi, and Alice Othmani. Mfcc-based recurrent neural network for automatic clinical depression recognition and assessment from speech, 2020.

[80] Hao Sun, Hongyi Wang, Jiaqing Liu, Yen-Wei Chen, and Lanfen Lin. Cubemlp: An mlp-based model for multimodal sentiment analysis and depression estimation, 2022.

[81] Joshua Y. Kim, Greyson Y. Kim, and Kalina Yacef. Detecting depression in dyadic conversations with multimodal narratives and visualizations, 2020.

[82] Michel Valstar, Jonathan Gratch, Bjorn Schuller, Fabien Ringeval, Denis Lalanne, Mercedes Torres Torres, Stefan Scherer, Guiota Stratou, Roddy Cowie, and Maja Pantic. Avec 2016 - depression, mood, and emotion recognition workshop and challenge, 2016.

[83] Pınar Baki. A multimodal approach for automatic mania assessment in bipolar disorder, 2021.

[84] Zahraa Al Sahili, Ioannis Patras, and Matthew Purver. Multimodal machine learning in mental health: A survey of data, algorithms, and challenges. *arXiv preprint arXiv:2407.16804*, 2024.

[85] Silvio Amir, Glen Coppersmith, Paula Carvalho, Mário J. Silva, and Byron C. Wallace. Quantifying mental health from social media with neural user embeddings, 2017.

[86] Anshu Malhotra and Rajni Jindal. Multimodal deep learning based framework for detecting depression and suicidal behaviour by affective analysis of social media posts. *EAI Endorsed Trans. Pervasive Health Technol.*, 6(21):e1, 2020.

[87] Palash Moon and Pushpak Bhattacharyya. Multimodal depression detection: A survey.

[88] Yanrong Guo, Chenyang Zhu, Shijie Hao, and Richang Hong. A topic-attentive transformer-based model for multimodal depression detection, 2022.

[89] Shweta Yadav, Jainish Chauhan, Joy Prakash Sain, Krishnaprasad Thirunarayan, Amit Sheth, and Jeremiah Schumm. Identifying depressive symptoms from tweets: Figurative language enabled multitask learning framework, 2020.

[90] Riccardo Franceschini, Enrico Fini, Cigdem Beyan, Alessandro Conti, Federica Arrigoni, and Elisa Ricci. Multimodal emotion recognition with modality-pairwise unsupervised contrastive loss, 2022.

[91] Soheil Khorram, Mimansa Jaiswal, John Gideon, Melvin McInnis, and Emily Mower Provost. The priori emotion dataset: Linking mood to emotion detected in-the-wild, 2018.

[92] Andrew G. Reece, Andrew J. Reagan, Katharina L. M. Lix, Peter Sheridan Dodds, Christopher M. Danforth, and Ellen J. Langer. Forecasting the onset and course of mental illness with twitter data, 2016.

[93] Neha Sharma and Kairit Sirts. Context is important in depressive language: A study of the interaction between the sentiments and linguistic markers in reddit discussions, 2024.

[94] Nawshad Farruque, Chenyang Huang, Osmar Zaiane, and Randy Goebel. Basic and depression specific emotion identification in tweets: Multi-label classification experiments, 2021.

[95] Virginia Morini, Salvatore Citraro, Elena Sajno, Maria Sansoni, Giuseppe Riva, Massimo Stella, and Giulio Rossetti. Who can help me? reconstructing users' psychological journeys in depression-related social media interactions, 2023.

[96] Alex-Răzvan Ispas, Théo Deschamps-Berger, and Laurence Devillers. A multi-task, multi-modal approach for predicting categorical and dimensional emotions, 2023.

[97] Andrea Laguna and Oscar Araque. A cost-aware study of depression language on social media using topic and affect contextualization, 2023.

[98] Jingjing Yang, Haifeng Lu, Chengming Li, Xiping Hu, and Bin Hu. Data augmentation for depression detection using skeleton-based gait information, 2022.

[99] Kunyao Lan, Cong Ming, Binwei Yao, Lu Chen, and Mengyue Wu. Towards reliable and empathetic depression-diagnosis-oriented chats, 2024.

[100] Shanya Sharma and Manan Dey. Assessing viewer's mental health by detecting depression in youtube videos, 2020.

[101] Sebastian Rodriguez, Sri Harsha Dumpala, Katerina Dikaios, Sheri Rempel, Rudolf Uher, and Sageev Oore. Predicting individual depression symptoms from acoustic features during speech, 2024.

[102] Ilija Tavchioski, Marko Robnik-Šikonja, and Senja Pollak. Detection of depression on social networks using transformers and ensembles, 2023.

[103] Yue Gu, Shuhong Chen, and Ivan Marsic. Deep multimodal learning for emotion recognition in spoken language, 2018.

[104] Juan Vazquez-Rodriguez, Grégoire Lefebvre, Julien Cumin, and James L Crowley. Emotion recognition with pre-trained transformers using multimodal signals, 2022.

[105] Pratiksha Meshram and Radha Krishna Rambola. Retracted: Diagnosis of depression level using multimodal approaches using deep learning techniques with multiple selective features. *Expert Systems*, 40(4):e12933, 2023.

[106] Qingkun Deng, Saturnino Luz, and Sofia de la Fuente Garcia. Hierarchical attention interpretation: an interpretable speech-level transformer for bi-modal depression detection, 2023.

[107] Sudhir Kumar Suman, Hrithwik Shalu, Lakshya A Agrawal, Archit Agrawal, and Juned Kadiwala. A novel sentiment analysis engine for preliminary depression status estimation on social media, 2020.

[108] Puneet Kumar, Alexander Vedernikov, and Xiaobai Li. Measuring non-typical emotions for mental health: A survey of computational approaches. *arXiv preprint arXiv:2403.08824*, 2024.

[109] Yuchen Cao, Jianglai Dai, Zhongyan Wang, Yeyubei Zhang, Xiaorui Shen, Yunchong Liu, and Yexin Tian. Machine learning approaches for mental illness detection on social media: A systematic review of biases and methodological challenges, 2025.

[110] Danielle Mowery, Craig Bryan, and Mike Conway. Feature studies to inform the classification of depressive symptoms from twitter data for population health, 2017.

[111] Thong Nguyen, Andrew Yates, Ayah Zirikly, Bart Desmet, and Arman Cohan. Improving the generalizability of depression detection by leveraging clinical questionnaires, 2022.

[112] David Owen, Jose Camacho Collados, and Luis Espinosa-Anke. Towards preemptive detection of depression and anxiety in twitter, 2020.

[113] Lin Sze Khoo, Mei Kuan Lim, Chun Yong Chong, and Roisin McNaney. Machine learning for multimodal mental health detection: a systematic review of passive sensing approaches. *Sensors*, 24(2):348, 2024.

[114] Yong Shan, Jinchao Zhang, Zekang Li, Yang Feng, and Jie Zhou. Mental health assessment for the chatbots, 2022.

AI-generated, for reference only.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.