Journal of Integrated and Advanced Engineering (JIAE)
Vol. 3, No. 1, March 2023: 47-64
http://asasijournal.id/index.php/jiae
http://doi.org/10.51662/jiae.v3i2.85

Conversational Analysis Agents for Depression Detection: A Systematic Review

Akeem Olowolayemo[1*], Maymuna Gulfam Tanni[1], Intiser Ahmed Emon[1], Umayma Ahhmed[1], 'Arisya Mohd Dzahier[1], Md Rounak Safin[2], and Nusrat Zahan Nisha[3].
[1]Deptartment of Computer Science, KICT, International Islamic University Malaysia, Malaysia
[2]Deptartment of Mechanical Engineering, Universiti Teknologi Malaysia, Malaysia
[3]Deptartment of computer science, University Malaysia Pahang, Malaysia

Abstract
Depression is known as a non-cognitive disturbance that can be seen among different people all over the world. This pertains to disorders that have affected cognitions and behaviors that arise from overt disorders in cerebral function. It is more common for young adults to elderly people based on lifestyles, work pressure, personal problems, diseases, people who had strokes or hemorrhages, certain brain diseases, and paralysis. This paper is focused on reviewing the research papers previously done on detecting depression. Utilizing predefined search systems, we have gone through a couple of studies zeroing in on gloom and involved conversational information for location and conclusion. The objective of this research is to review large research studies on whether conversational agents can detect and diagnose depression by using smart texting analysis. The study was done by searching IEEE Xplore, Sci-hub, Doi, Scopus, and Pubmed using a predefined search strategy. This review was focused on studies that include the possibilities and steps of detecting depression and diagnosis that involved conversational data or analysis agents after assessing them by independent reviewers and relevancy for eligibility. After retrieving more than 117 references initially it was narrowed down to 95 references that were found relevant as most of them applied analytical techniques and technology-based solutions. Detecting depression and diagnosing it through smart texting analysis is a broad and emerging field and has a promising future but not every research studies were robust enough to get valid results in the end. This study aimed to keep the review as precise and informative as possible.

Keywords:
Artificial Neural Network;
Conversational Agent;
Depression;
Mental Illness;

Article History:
Received: August 22, 2022
Revised: February 21, 2023
Accepted: February 27, 2023
Published: March 9, 2023

Corresponding Author:
Akeem Olowolayemo
Deptartment of Computer Science, KICT, International Islamic University Malaysia,
Deptartment of Mechanical Engineering, Universiti Teknologi Malaysia, Malaysia
Email: akeem@iium.edu.my

INTRODUCTION

The growing number of people dealing with depression cannot be ignored which can be challenging in the social and economic state of any country. Depression usually can be seen among people of all ages. In most cases, depression cannot be fully treated with medicine [1]. The other kinds of diseases might have devastating consequences, but these kinds of neuropsychiatric disorders, depression are something people must deal with for a very long time. Nowadays youngsters and sometimes elderly people may go through clinical depression and other mental health issues which most of the time are ignored compared to health issues such as depression, insomnia, etc. which can lead them to other age-associated physical diseases. These mental health issues should not be only treated with medicines but also with therapy, religious and spiritual activities, and support from family and children [2]. These kinds of consultation and therapy are now being provided by e-health which has become much more advanced over the past few years [3]. Technologies and healthcare have built a bridge between them to improve our lifestyle with the help of a vast range of medical and clinical datasets [4].

Now artificial intelligence is being used to take this bridge to a high standard in decision making from surgeries to consultation [5].

As computerized reasoning can emulate the way a human can think, that is the reason it was the primary technique created [6] which is propelled by the dynamic cycle in the organization of natural nerve cells by changing contributions with the connection of the organization to deliver a result. ANNs can learn the situation and make decisions accordingly [7]. From the study in [8], the conversational agents were introduced to the emerging health-related research field for more stable experimental designs. Nowadays we can see depression is now treated or diagnosed by medical experts such as psychiatrists or neurologists for more stable and controlled treatment but the diagnostic tools that have been used for collecting neuropsychiatric scales such as MMSE for dementia [9][10], PHQ-9 for depression [11] as these tests provide result based on the mood-related state of a patient to evaluate.

There are a few disadvantages of the current diagnosis process as sometimes it can be confusing even for the doctors to detect characteristics [12] of dementia [13] or depression [14] as in most cases patients rely on taking medical help at a very late stage of these disorders and mental conditions. For this reason, it can be a bit challenging for doctors to differentiate and treat them.

The aim was to identify research proposals that are already existing in the use of conversational systems for detecting and diagnosing clinical depression, and neuropsychiatric disorders based on artificial neural networks. Scientific literature was surveyed to evaluate studies on detecting depression by using conversational agents. The analysis provides an overview of how specific disorders and symptoms of depression are being identified and what kind of technologies can be used for collecting data to enhance performance and detect the gaps in the existing research and build a bridge in these relevant research papers. As clinical depression is a broad field to work on, there are so many research studies that have been already published. Reviewing many studies, this review paper tried to be precise as well as informative, to highlight the parts that will collect data and using the help of technology for effective detection of depression, and it can be gradually diagnosed. From this study, gradual progress has been made in this field with the help of artificial intelligence. The study aims to provide an insight into how much further research has to be done and how these research studies can bring revolutionary changes. It equally underscores where some of them have been inadequate or failed to explain valid outcomes has also been explained in this review paper briefly. The structure of this paper will follow methodology, analysis of results, discussion, and conclusion.

METHODS

The examination procedure was performed following the preference of PRISMA. A convention for directing an orderly writing survey and meta-examination is characterized by PRISMA. The meaning of this convention is a progression of undertakings that cover the survey strategy, beginning with the preparation of the survey to the selection of sources of data, to the real meta-investigation of specific works. PRISMA flowchart is presented in Figure 1.

Search strategy

The goal of the pursuit cycle was to find text examination arrangements regarding depression, as such, recommendations that depend on regular language, and text investigation, as a connection component. This generally incorporates, yet is not restricted to, excellent AI-based visiting applications, which is the innovation fundamentally designated from our exploration.

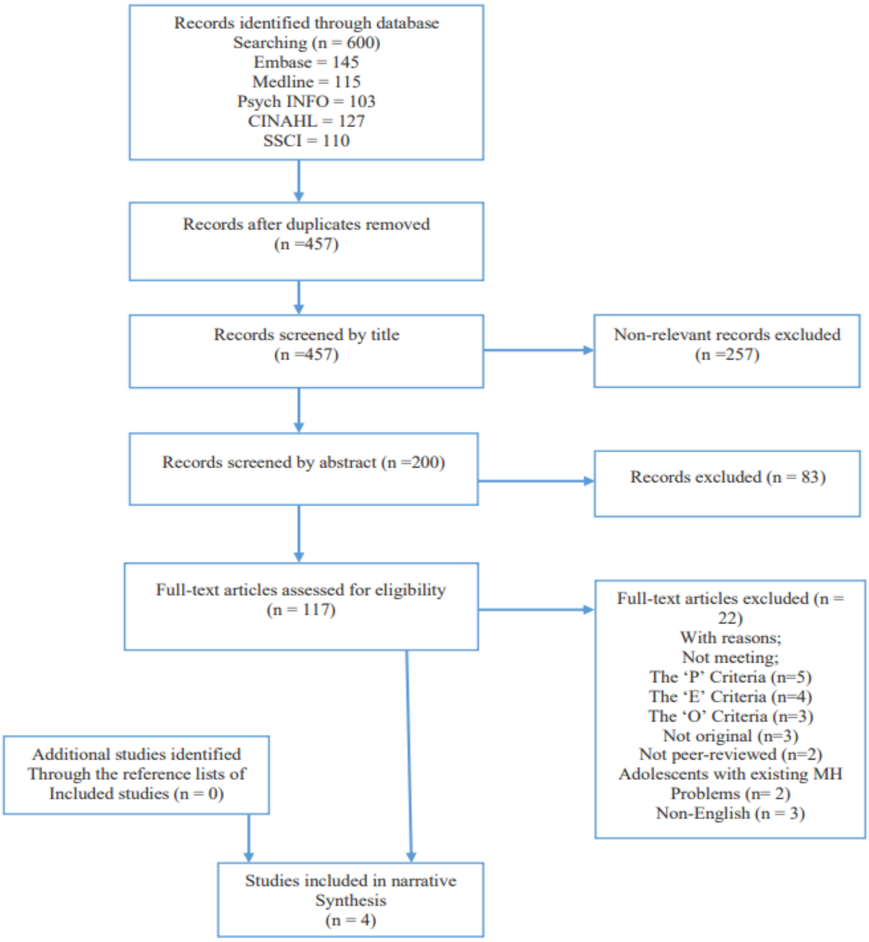

Figure. 1 PRISMA flowchart

Thus, any study that aligns with conversational innovation of any sort has been explored, including chatbots. The consideration of commitments addressing or connecting with the examination of psychological illnesses, (for example, melancholy, dementia, self-destructive contemplations, and so on) as well as other huge issues that the populace experiences in the current period, like social seclusion, was one of the determination standards. To examine the previously mentioned issues or in some alternate manner survey the client's state regarding these issues, the recommendations ought to incorporate an examination of the conversational information. Articles that do not meet any of the dismissal rules will then, at that point, be also filtered in the accompanying stage by elimination screening.

Eligibility criteria

In this research, we examined a wide range of articles and papers. Based on the eligibility criteria, we can summarize the dismissal rules for all the articles as follows. Not every publication is about depression, while a few publications place more emphasis on social media use. Some other studies do not demonstrate how to identify depression using text analysis. In some other studies, the analysis was insufficient to demonstrate that people who experience sadness or any other issues based on the conclusions. Some papers do not contain appropriate information. As many papers have been chosen, some of these are not related to the aim of this systematic literature review. Hence, we have only reviewed studies that are closely related to our topic of study which means we have only explored papers that are more focused on how to detect depression from conversational agents.

Even if some of the publications reviewed do not directly relate to them, they do contain some information on mental illnesses. Particularly, several publications concentrate on the use of questions to determine whether a person has a mental disorder.

Selection process

Mendeley functionalities were utilized to acquire all the references from all informational collections addressed. Then, at that point, we applied a matching basic instrument to recognize duplicates and eliminate them. Generally, continuous disseminated papers were focused. Also, attempts were made to track down related papers on Google scholar. We adhere strictly to picking suitable resources and followed a combination of strategies to pull together every one of the studies related to distinguishing mental problems through text examination.

In the principal screening stage, we completed an underlying determination of words given their significance regarding the downturn and demonstrations of worrying sentences. During this screening, words, and sentences were utilized to recognize assuming any of the materials included any messages or content that can relate to wretchedness. The screening is finished by the qualification measures referenced before. Toward the end, the suitable references meeting the set criteria have been utilized in the review study.

RESULT AND DISCUSSION

The goal of this systematic review is to get the outcome from desired research presented by investigating important themes concerning conversational frameworks for the detection of neuropsychiatric issues, such as depression. This was performed achieved by following the PRISMA procedure [15], to gather, aggregate, dissect and examine the principal discoveries of target papers.

As portrayed in the PRISMA flow diagram, a total of 600 papers were extracted from all the online sources. After exclusion based on the eligibility criteria, 457 papers in addition to 3 extra articles distinguished in the collected papers were reviewed, from topics as well as modified works, to identify those articles that were possibly pertinent to be included in the exploration set. Ultimately, 117 papers have been named relevant considering topics and uniqueness.

Then, the qualification stage was conducted on the 117 chosen papers and each of them was checked properly. We divided the works into three groups, specifically the papers that were significant and directly related to the survey for additional examination. Some papers were disposed of as those papers were considered non-important, and the pertinent papers were checked once again by the researchers. Finally, the two researchers completed their procedure to check all the papers and agreed to choose some relevant papers in the third stage.

From the selected 117 articles followed by the eligibility stage, 12 papers were chosen as applicable after perusing the whole topic, though 25 papers were considered indistinct. Finally, after a conversation of the writers taking part simultaneously, 5 were at last chosen for investigation. The leftover 22 (14 muddled and 8 non-significant) were disposed of demonstrating their comparing prohibition rules. The principal justification for the avoidance of the 15 articles was the way that the arrangement talked about was not conversational communication utilizing regular language whether it is spoken or typed, inside the framework. Additionally, 13 papers were disposed of due to their irrelevance to the topic as these topics were not related to mental health or neuropsychology. After that, 8 papers were disposed of due to the study that was not centered on recognition or conclusion of such issues, instead counteraction, intercession, help or treatment, or a blend of them. Another two papers were disposed of because their discovery instrument depended on the utilization of information inconsequential to the discussions between the client and the framework, for instance, face demeanors or actual area and movement indicators. After the selection of the excess papers, 10 full-text archives have been removed in all the ways accessible to the writers. At last, 7 papers were audits, studies, or gatherings of meeting procedures. Even though these articles were explored for possibly pertinent references, they were excluded from this review.

Toward the finish of the cycle, 4 articles were chosen to be entirely examined by the researchers, with the primary goal to separate data around: (1) the particular goal or goals of each review; (2) the particular neuropsychology or mental problem that the recognition framework is engaged on; (3) the nature of the clients which the conversational system is centered around, and particularly their age; (4) the progressions used for encouraging the conversational system, UI or data assessment; (5) relevant information procurement model which is the manner in which the framework changes and gathers conversational information in regular words and must be perfect to handle the examination component; (6) designs supporting the identification for making expectations in light of the information gained from discussions of the topic; and lastly, (7) the approval convention conducted, with the highest quality level checking utilized and also the quantity of genuine clients partaking in it. The improvement of AI and its rising application to a variety of areas, including emotional well-being, has carried with it the need to examine and control this application morally. In this way, we are currently at another convergence point, where the blend of AI and emotional well-being raises the need for clever examinations [16]. For several years now, telehealth administrations, particularly portable and cell phone applications for emotional well-being, have been deeply grounded in the focused field. A telerehabilitation study was the most common among the studies. These outcomes mirror the developing examination patterns in e-psychological wellness care [17].

Discussion

According to several studies, three key goals are primarily the focus of studies on conversational analysis for the diagnosis of mental illnesses [18]. The three primary targets are (i) identification based on language and verbal information, word choices, and punctuation [18]; (ii) discovery based on voice manner and tone; and (iii) location based on non-verbal communication appearance while transmitting thoughts or the individual's look. Additionally, body language has been considered an important feature for detection. Each approach has different conditions. When more than two people are communicating, using language, or chatting on social media, the first method of detection is conceivable. The second strategy can be used when speaking or sending voice notes. The last strategy is appropriate if someone can see a person's body language [19]. For people suffering from depression or frustration, some non-verbal changes can be noticed [20]. It fluctuates depending on the person. Less eye contact and slouching can be noticed depending on the individual. Their hand movements could turn out to be slower or less successive. To gather verbal conversational data, most of them [18,19,20,21] use an AI-driven Conversational Agent (CA). The message is used to process the most suitable framework reaction and to separate the semantics of the message being sent [22]. The text additionally uncovers insights regarding the user's close mental health condition. This is passed on through language and linguistic development [22]. Discussions among patients and nervous system specialists might uncover correspondence issues, as per late exploration utilizing the subjective approach of Conversational Agent (CA) [21], and this can be utilized to recognize patients with neurodegenerative problems (ND) and those with functional memory disorder (FMD), displaying memory issues insignificant to dementia [21, 26, 27]. To identify dementia unusual questions can be asked. The essential conclusions of these studies are that individuals utilize psychological sickness and sadness-related phrases in various settings and that material containing words connected with self-destruction and sorrow are simply extraneously related to genuine self-destructive ideation or potentially depression [32]. The study in [33] explains why there has recently been a considerable need for early biomarkers of mild mental decline in the context of Alzheimer's Disease (AD) and dementia research. The focus of the study in [34] is on developing a chatbot that can lead a meeting based on the Patient Health Questionnaire (PHQ) -9 exam to increase screening for openness, allure, and the possibility of melancholy. Particularly linguistic patterns were discovered to be linked to depression [35]. So, the review proves that individuals' language is a decent mark of their mental states, specifically, linguistic patterns were discovered to be linked to depression [35]. Many people who experienced stress find it simpler to communicate their emotions online. Different emotions can be added to the sentiment scores to improve the process of calculating depression ratings [36]. This procedure highlights several unappreciated characteristics of depression. The process of extracting feelings and views from user-posted tweets is known as sentiment analysis of Twitter data. At least 92 percent of youngsters using social media, suffering from depression can be determined by linguistic features of a written text and the level of current depressive symptoms is worrying. To sum up, it is suggested that depression can be predicted from a casual language composed using quantitative linguistic features that are best suited to the methodology.

Detecting suicidal and depressional disorders:

How conversational analysis can prove someone is suffering from depressional disorders and possible suicidal ideation is discussed here. The present application of the text-based technique for detecting early depression is observed in this systematic review. The American Psychiatric Association's (APA (2013)) latest update of the scientific categorization and indicative manual, the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), shapes the reason for the recommended grouping. It uses commonly utilized phrasing on guidelines [35]. Anyone who experiences major sadness for most of the day nearly every day for several weeks or months may have clinical depression. People who are depressed could lose interest in almost all activities, feel exhausted or stressed, or have trouble focusing or remembering things. These symptoms are often expressed while posting on social media. Self-destruction might seem like a way to reestablish respect and control of one's life if one is experiencing an ongoing condition without really any possibility of treatment or help from the desolation. This study also provided a summary of the approaches currently being used for conversational or text-based depression detection while some equally involve voice analysis. In conclusion, the articles reviewed focus on mental diseases. Some other pressing issue in this category especially among elderly patients is dementia, while specific manifestations like Alzheimer's Disease (AD) or its intermediate condition moderate cognitive impairment (MCI) are also addressed [35]. On the other hand, depressive disorders are the focus of a sizable number of other publications, which collectively encompass a sizable portion of mental disorders. Main symptoms to identify mental disorders is listed in Table 1. Table 2 show some features designed for acoustic signal emotion detection.

Table 1. Main symptoms to identify mental disorders

Objectives	Sub-objectives	References
Detection based on content/message analysis	AI-driven Conversational Agent	[18,21]
Detection based on vocal voice analysis	AI-driven Conversational Agent	[18,20, 21]
Visual expression and facial expression	Body languages	[18,19,20]

Table 2. Features designed for acoustic signal emotion detection [22,23,24,25]

Groups	Characteristics	Physiological modifications brought on by emotion
Key	The least and most elevated esteems, the mean, median, standard deviation, the worth in the first and last vocal fragments, the relationship coefficient, the slant, and the straight relapse blunder.	The subglottal air pressure and the vocal fold tension.
The initial two formant frequencies' data transfer capacities	The esteem in the first and only articulated pieces as well as the reach, mean, middle, and standard deviation.	Voice boxes reverberate.
Energy	Energy, alongside the upsides of the underlying and finishing up voiced fragments, relationship, incline, and mistake, is the negligible worth of the energy direct relapse.	Vocal voice effort. Emotional arousal.
Pattern	Speech rate, the proportion of voiced to unvoiced segments, the size of the largest voiced segment, and the lengths of both voiced and unvoiced segments.	Conditions are troublesome and distressing to emotional wellness.

Table 3. Text-based depression detection techniques

Methods	References
Scikit-Learn Toolkit machine classification	[38]
Support Vector Machine	[38,39,40]
Classifiers	[38,39,40]
Clustering Approach	[41]
Logistic	[42]
BiLSTM + Attention	[43]
Gaussian Process	[42]

Table 3 presents the techniques of the strategies utilized in the text-based approach for distinguishing any mental health issue. The most frequently involved techniques in the arrangement of chosen articles are support vector machines (SVM). The BiLSTM + Attention model delivers the best results on depression from text analysis. This might be overgeneralizing, however, given the datasets utilized and the issues being tended to contrast.

Technological solutions applied

In this section, we analyze the technology platforms used in the reviewed works. In terms of the solution mentioned, studies in [46][59] and [63] were developed for an Android smartphone operating system or tablet, while [28] is also available for iOS devices. The study in [58] uses an iPad tablet as a frontend device, which runs iOS, and it can be also run in MacOS as for the Apple Silicon. Only two studies use the Cloud as a primary platform reference [53] and [56] uses Google Cloud Functions and Firebase. In [61] and [64] can be used in any HTML 5-based web browser and windows-based solutions in [33,61,62].

The research by [48] and [53] utilized TensorFlow innovation for profound learning-based examination as well as picture acknowledgment. TensorFlow is an end-to-end open source platform for machine learning with a comprehensive, flexible ecosystem of tools, libraries, and community resources. TensorFlow innovation is utilized in references [48] and [53] for profound learning-based examination and picture acknowledgment, in both. TensorFlow gives a work process system for developing and preparing models in Python or JS, while likewise effectively conveying them in the cloud. Table 4 list a synthesis of several studies. Also, the technologies used is shown in Table 5.

Table 4. Synthesis of several studies

Year	Type	Title	Objective	Scope	Target users	Conclusions	Technology	Data acquisition model	Analysis	Gold Standard	Validation
2022	Journal Article	Deep Learning for Depression Detection from Textual Data [45]	Identifying depression on social media based on text.	Depression	Any	Implementing the suggested method resulted in a lowered false positive rate and accuracy of 98%.	SW: TensorFlow, Keras library	RNN approach and the LSTM technique.	Principal Component Analysis (PCA)	RNN and LSTM	More than 4000 tweets.
2019	Conference Proceedings	Virtual Human Questionnaire for Analysis of Depression, Anxiety, and Personality [46]	Compare the equivalent of a virtual human interviewer and a virtual human agent.	Depression	Any	Individuals' reactions to surveys while being addressed by a virtual human versus when self-control similar show no perceptible contrasts.	SW: Aria-Valuspa Platform	Sound examination modules incorporate discourse acknowledgment and discourse action and turn-taking.	Manual	BFI-10, PHQ-9 and GAD-7	55 participants

Table 4. Synthesis of several studies (continued)

Year	Type	Title	Objective	Scope	Target users	Conclusions	Technology	Data acquisition model	Analysis	Gold Standard	Validation
2017	Journal Article	Combining speech-based and linguistic classifiers to recognize emotion in user spoken utterances [22]	Identifying emotion from spoken utterances by combining linguistic-based variables with acoustic and contextual aspects.	Emotion identification	Any	The results show that the combined results exceed the individual hypotheses and shed light on the features and classifiers available at each step, including detection and fusion.	SW: WEKA Toolkit, Apache OpenNLP	ASR System and algorithmic feature extraction	MLP for categorization based on sound. A few ML categories are based on text. Fusion technique based on voting.	Dataset with annotations.	3 datasets, totaling 482 dialogs.
2017	Journal Article	Virtual Human as a New Diagnostic Tool, a Proof-of-Concept Study in the Field of Major Depressive Disorders [47]	Evaluate the effectiveness of a major depressive disorder diagnosis approach that relies on an ECA's recognition of symptoms in outpatients.	Depression	Any	93% of specificity for the ROC analysis. Reached 73% with severe depressive symptoms. Acceptability of the ECA score was high (25.4)	SW: SPSS, Unity, MedCalc HW: 40' Display, Windows Desktop, Kinnect Sensor	voice recognition software	Classifier using a decision tree	MDD DSM 5	179 outpatients, 35 diagnosed with MDD

Table 5. Technologies Used

Technology Platforms	References
MacOS	[58,28]
Android	[46,59,28,63]
Linux	[71,65]
Windows	[33,61,62]
WebApp	[61,64]
IOS	[58,28]
Google Cloud	[53,56]
TensorFlow	[48,53]
Not accessible	[46,48,49,50,51,34,57,67,68]

Summarily, with regards to technology platforms, cell phones, regularly Android-based, are utilized in most of the surveyed works because of their convenience and moderateness. Interestingly, demanding innovations are huge, and the evaluated works adopted dissimilar ones. Lastly, countless articles either require no exceptional necessities to run or do not indicate which ones were used. The absence of consistent technology platforms shows that this is a moderately new field of study and most of the studies evaluated are in the exploratory stage, using various specialized conceivable outcomes.

Data acquisition and analysis

The methodology and algorithms used to process and analyze conversational data are described in this section. These comprise characteristics and other information collected from speech or text sequences. The goal of data gathering and analysis is to finally understand the mental health of the target user. We built a classification based on the data source, and this section describes conversational data, which contains characteristics and other information collected from speech or text sequences. We developed a classification based on the data source, and as a result, each system may have multiple sources. Table 6 summarizes the data acquisition procedures, which include conversational and environmental data that were prepared for analysis.

The very first data source is verbal statements and language. The audio from the microphone is processed, either instantly or offline, and statements, as well as all associated features, are extracted. This can be done manually or automatically with the help of software tools. In [69,71,72,75,76,77,79,81,84], automatic speech recognition is used, whereas transcription is done offline in [68,73,80], and [74]. Moreover, the study in [75] supports text-based interaction, which eliminates the need for Automatic Speech Recognition to extract verbal information.

Table 6. Data acquisition procedures

Procedure		References
Speech	Real-time transcription	[69,71,72,75,76,77,79,81,84]
	Offline transcription	[68,73,74,80]
Audio-visual	Automatic feature extraction	[49–73, 76,77,78,80,82,83, 84]
Environment sensors	Automatic feature extraction	[72, 78]

Table 7. Analyzing methodologies

Method	Algorithm	References
Machine Learning	Logistic regression	[50,33,59,61]
	SVM	[51, 33,34,57,59,61,64]
	Decision Tree	[52,59,62, 64]
	Other	[46,57, 59]
Deep Learning	MLP	[43,56, 33]
	Other	[48,53,34,56,57]

Audios and videos collected during sessions and discussions are other sources of data. However, instead of extracting words, they are now used to directly extract additional attributes, such as voice pitch, speech pace, or face expression features. The work in [49–73, 76,77,78,80,82,83,84] employ audio features, whereas [71,72, 74, 78] employ both audio and video elements. In all cases, feature extraction is done automatically.

Concluding with analyzing methodologies, additional features derived from various environmental sensors are included in Table 6. A GPS monitor, object identification functions, activity recognition, and other sources such as temperature and health sensors are used [53]. The study in [59] on the other hand, captures indoor location activity using mobile transmitters. Table 7 shows an analyzing of several methodologies.

In terms of analysis algorithms and methods, Machine Learning (ML) is used in most of the studies [46,57, 59]. Machine Learning is a branch of Artificial Intelligence (AI) that studies how computer algorithms can improve themselves automatically over time [26], typically based on mathematical algorithms that "learn" by adjusting to patterns in a training dataset. If new data is provided, the previously trained algorithm will be able to recognize it. The studies in [51, 33,34,57,59,61,64] use Support Vector Machine (SVM), [52,59,62, 64] use Decision Trees, although in [52] and [64] a combination of Decision Trees called Random Forest is used. Additionally, [46, 57,59] use several algorithms and methods, [52] uses Linear Discriminant Analysis (LDA), [50] uses Linear Regression, and [46] relies on the Naïve Bayes and Maximum Entropy (MaxEnt) algorithms.

Deep Learning (DL), which is technically part of ML, is used in an even greater number of articles, included in a distinct category because it is a very new and flourishing discipline with an entirely different approach to feature extraction than traditional ML. Deep Learning is an area of Machine Learning algorithms that incorporates many processing layers. They enable the simultaneous processing of multi-level forms required for feature extraction and classification from raw data, performed by putting together a collection of simple but non-linear modules that convert a representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level [46].

A Multilayer Perceptron (MLP) is used in references [43,56, and 33]. A Hierarchical Collective Agent Network (HCAN) is used in [48] while Probabilistic Neural Networks (PNN) and Extreme Learning Machines (ELM) are also utilized in [64]. Moreover, [56] employs Reinforcement Learning, a sort of machine learning that does not require labeled data for training — instead, it can learn by interacting with the system and analyzing its response. is used not just for detection but also for modeling the dialogue as a Markov Decision Process. A multidimensional fusion method will be used in [24], but it is not specified, as it is in [57].

Finally, no machine/deep learning algorithm is used in refs. [52,60], [65], or [67], as the analysis is done manually. In conclusion, regardless of the data source, an enormous number of papers use automatic feature extraction. Then, one or more machine learning algorithms (SVM is particularly common) are used to produce predictions about the user's mental health. Furthermore, we've found that more complex models, such as Deep Learning and Reinforcement Learning, are gaining popularity as a result, AI plays an important role in the examined works, hinting that this field of study would dominate in future improvements in health care applied to depression and suicide detection.

Validation

In this study, the Standards for Educational and Psychological Testing [48] were used as a reference. These guidelines specify reliability and validity as the two key characteristics of any test. Validity focuses on a method's precision and correctness, whereas reliability assesses a method's consistency and dependability. Only two of the analyzed works compute some form of metric in terms of reliability [71,79]. Cronbach's Alpha coefficient is derived using pairwise correlations between items to assess dependability [48].

Regarding the validity component, only the article's "criterion validity"—which assesses how closely the suggested test's results match those of an alternative, valid test—was considered. To measure this, precision and recall values, two extensively used metrics in the field of statistical analysis of binary classifications, are often used. They are also combined with additional metrics derived from them, such as the F1-score and the Area Under the Receiver Operating Characteristic. To measure this, precision and recall values, two extensively used metrics in the field of statistical analysis of binary classifications, are often used. They are also combined with additional metrics derived from them, such as the F1-score and the Area Under the Receiver Operating Characteristic. Most of these works rely on a gold standard test to perform validation. A gold standard is a reference test that has been previously validated and is used for comparison with a novel method. The most often used gold standard tests by the examined articles are mentioned in the Validation table. The Patient Health Questions is a self-report depression evaluation that assigns a score based on several DSM-IV criteria [13]. The PHQ-9 version is used as the gold standard in references [48,71], [75], and [79], while the PHQ-8 version is used in references [48]. The ninth item of the PHQ-9, which is used to gauge the risk of suicide, asks about both passive and active thoughts of harming oneself [83]. This makes it different from the other versions of the questionnaire.

The Beck Depression Inventory-II is another widely utilized gold standard exam for determining the severity of depression (BDI-II). The BDI-II has 21 questions, and like the PHQ, each response is graded on a scale from 0 to 3. More severe depression symptoms are indicated by higher total scores [72]. [78] and [82] use the BDI-II as their benchmark. Table 8 and Table 9 show the validation standards, attributes and metrics, respectively.

The Mini-mental State Examination is by far the most common memory test (MMSE). It is widely used to assess cognitive problems in clinical and research contexts [12]. It assesses time-space orientation, (ii) attention, (iii) memory, (iv) language, and (v) complex instructions and can award up to 30 points. The MMSE test is referred to be the gold standard in references [75, 82,81] and [73]. References [84] and [81] make use of already-existing patient datasets that each contain information on a patient's reference mental health status.

Table 8. Validation Standards

Gold Standard	Category	References
Patient Health Questions (PHQ)	Depression	[48, 71, 75, 79]
Beck Depression Inventory-II (BDI-II)	Depression	[81,83]
Mini-mental State Examination (MMSE)	Suicidal	[72,74,77,80]
Database with diagnosis	Depression	[69,66]
Other	Suicide	[70, 71, 72, 74, 77,78, 82]

Table 9. Validation attributes and metrics

Validation attributes	Metrics	References
Reliability	Cronbach's Alpha	[75, 81]
Validity	Precision and recall	[68, 69, 70, 73, 74, 75,76, 78, 79, 81, 82]
	Accuracy	
	F1-score	
Usability	AUROCH	[84]
	System Usability Scale (SUS)	
Acceptability	E-scale (AES)	[81]

Six references, including [68, 69, 70, 73, 74, 75,76], [78], and [82] use less common alternative gold standard tests, with [22] and [26] being unique in that they use clinical experts' diagnoses as their reference. The user experience was also measured by tests in two references. The study in [68] employs the Acceptability E-scale (AES) [70] to evaluate the acceptability of the Embodied Conversational Agent in use, whereas [84] evaluates the product's perceived system satisfaction and usability using the System Usability Scale [53].

In the end, the validation of the works is somewhat sporadic and insufficient, as is to be expected in such a fresh subject. Even though many of them rely on known gold-standard tests, only two works do a reliability study and just two compute any kind of user experience metrics (especially the PHQ-9 for depression and the MMSE for dementia). Furthermore, criterion validity is the only valid type that is measured. To support the adoption of these revolutionary procedures in the therapeutic setting, extensive validity (facial, content, concept, criteria, internal and external validity) and reliability (test-retest, interobserver, and internal reliability) tests must be performed. To prove that these procedures can attain the same diagnosis capabilities, a validation process comparable to the ones used by the gold standard tests should be carried out. It is noteworthy that out of the 11 works that rely on real users testing their systems first-hand, 9 of them involved 54 or fewer participants. The statistical significance of the research generally ranges up to one order of magnitude.

This reinforces our belief that it is very challenging to conduct a proper validation process with actual participants given the novelty of the issue. Therefore, in subsequent research on this subject, this component should be improved.

Limitations

The discoveries of this systematic survey added to studies on the impacts of discouragement as well as other related results like inspirations for self-destructive endeavors in individuals who are bound to encounter tension and melancholy, yet they additionally offered significant proof for the adverse consequences of web-based entertainment use on psychological well-being. One of the most recognizable signs of depression is a person's facial expression. When someone consults a psychiatrist, the psychiatrist finds several signs that point to difficulties. For instance, understanding a person's body language, facial expressions, social activities, and health issues. To obtain more precise results, it is crucial to evaluate these. The findings of this research demonstrate that (1) Ethical contemplations, (2) Scarcity of information, and (3) Not having enough comprehension of emotional well-being are the three most concerning issues. One of the other earlier evaluations that this review's eligibility phase highlighted as a source of additional material to obtain a deeper insight into this area [93,94,95] was the usage of conversational data for health objectives.

CONCLUSION

To conclude, this study presents a systematic review of using conversation systems in the field of clinical depression detection and diagnosis. The main motivation was to find if it was possible to detect depression with the help of technologies analyzing different behavioral changes for addressing specific disorders based on many studies. The study found that depression can be detected through conversational agents in most cases. It can be depicted from analyzing the reviewed papers that there are two sides. On one side smart conversational agents for detecting depression are very promising in the upcoming years and on the other side, it seems quite challenging to detect behavioral changes as it has so many aspects and it can be confusing for any machines designed artificially to detect them correctly and more specifically medical standard-wise. For improved accuracy, looking at facial expressions is necessary. But over the next few years, AI will be more advanced, effective, reliable, and more stable so that medical experts such as doctors can more reliably detect depression from an early stage and diagnose it precisely.

REFERENCES

[1] M. Kowalska *et al.*, "Aging and Neurological Diseases," in *Senescence - Physiology or Pathology*, InTech, 2017, doi: 10.5772/intechopen.69499.
[2] J. Rodda, Z. Walker, and J. Carter, "Depression in older adults," *BMJ (Online)*, vol. 343, no. 7825. Oct. 01, 2011, doi: 10.1136/bmj.d5219.
[3] V. della Mea, "What is e-health (2): The death of telemedicine?," *Journal of Medical Internet Research*, vol. 3, no. 2. JMIR Publications Inc., pp. 6–7, 2001, doi: 10.2196/jmir.3.2.e22.
[4] Institute of Electrical and Electronics Engineers and IEEE Communications Society, *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC).*
[5] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," *Proposal*, Dartmouth University, 2006.
[6] NN, "*Introduction and Role of Artificial Neural Networks*," 2019. [Online]. Available: www.worldscientific.com
[7] M. Stankovic, V. Kovanovic, and J. Jovanovic, "Intelligent Software Agents and Multi-Agent Systems Feedback Actionability in Learning Environments View project GRASS-Grading Soft Skills View project," *Encyclopedia of Information Science and Technology Edition: 2ndPublisher: Idea Group International Publishing*, 2009.
[8] L. Laranjo *et al.*, "Conversational agents in healthcare: A systematic review," *Journal of the American Medical Informatics Association*, vol. 25, no. 9. Oxford University Press, pp. 1248–1258, Sep. 01, 2018. doi: 10.1093/jamia/ocy072.
[9] M. F. Folstein, S. E. Folstein, and P. R. Mchugh, "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician Related papers," *Pergamon Press*, 1975.
[10] T. Peitzker, "Why chatbots are so disruptive," 2016. [Online]. Available: https://venturebeat.com/2016/08/16/why-chatbots-are-so-disruptive/AIGUEST
[11] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9 Validity of a Brief Depression Severity Measure," *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001, doi: 10.1046/j.1525-1497.2001.016009606.x
[12] C. M. H. Hosman and E. J. Llopis, "Prevention of mental disorders: effective interventions and policy options: summary report CAMHEE-Child and Adolescent Mental Health in an Enlarged Europe View project Promising Neighbourhoods View project," *Summary Report*, World Health Organization, Geneva2004.
[13] D. M. Holtzman, J. C. Morris, and A. M. Goate, "Alzheimer's Disease: The Challenge of the Second Century," *Sci Transl Med*, vol. 6, no. 3, pp. 77, 2011, doi: 10.1126/scitranslmed.3002369
[14] M. Liaw, A. Wei, A. Olowolayemo, M. Gulfam Tanni, M. Abdullah Omar Ba Khadher, "Undergraduate Students' Mental Health Remote Monitoring System: End-User Usability Testing Study," 2022.
[15] A. A. Selcuk, "A Guide for Systematic Reviews: PRISMA," *Turkish Archives of Otorhinolaryngology*, vol. 57, no. 1, pp. 57–58, May 2019, doi: 10.5152/tao.2019.4058.
[16] S. D'Alfonso, "AI in mental health," *Current Opinion in Psychology*, vol. 36, pp. 112–117, Dec. 01, 2020. doi: 10.1016/j.copsyc.2020.04.005.
[17] T. Timakum, Q. Xie, and M. Song, "Analysis of E-mental health research: mapping the relationship between information technology and mental healthcare," *BMC Psychiatry*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12888-022-03713-9.
[18] M. R. Pacheco-Lorenzo, S. M. Valladares-Rodríguez, L. E. Anido-Rifón, and M. J. Fernández-Iglesias, "Smart conversational agents for the detection of neuropsychiatric disorders: A systematic review," *Journal of Biomedical Informatics*, vol. 113, 2021. doi: 10.1016/j.jbi.2020.103632.
[19] F. Loi, J. G. Vaidya, and S. Paradiso, "Recognition of emotion from body language among patients with unipolar depression," *Psychiatry Research*, vol. 209, no. 1, pp. 40–49, Aug. 2013, doi: 10.1016/j.psychres.2013.03.001.
[20] G. Schwartz, P. Salt, M. R. Mandel, and G. L. Klerman, "Facial Expression and Imagery in Depression: An Electromyographic Study." *Psychosom Med*, vol. 38, no. 5, pp. :337-347, 1976, doi: 10.1097/00006842-197609000-00006
[22] D. Griol, J. M. Molina, and Z. Callejas, "Combining speech-based and linguistic classifiers to recognize emotion in user spoken utterances," *Neurocomputing*, vol. 326–327, pp. 132–140, Jan. 2019, doi: 10.1016/j.neucom.2017.01.120.

[23] J. H. L. Hansen, “Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition,” vol. 20, no. 1-2, pp. 151-173, 1996, doi: 10.1016/S0167-6393(96)00050-7

[24] D. Ververidis and C. Kotropoulos, “Emotional speech recognition: Resources, features, and methods,” *Speech Communication*, vol. 48, no. 9, pp. 1162–1181, Sep. 2006, doi: 10.1016/j.specom.2006.04.003.

[25] D. Morrison, R. Wang, and L. C. de Silva, “Ensemble methods for spoken emotion recognition in call-centres,” *Speech Communication*, vol. 49, no. 2, pp. 98–112, Feb. 2007, doi: 10.1016/j.specom.2006.11.004.

[26] C. Elsey *et al.*, “Towards diagnostic conversational profiles of patients presenting with dementia or functional memory disorders to memory clinics,” *Patient Education and Counseling*, vol. 98, no. 9, pp. 1071–1077, Sep. 2015, doi: 10.1016/j.pec.2015.05.021.

[27] D. Jones *et al.*, “Conversational assessment in memory clinic encounters: Interactional profiling for differentiating dementia from functional memory disorders,” *Aging and Mental Health*, vol. 20, no. 5, pp. 500–509, May 2016, doi: 10.1080/13607863.2015.1021753.

[28] T. Ujiro *et al.*, “Detection of dementia from responses to atypical questions asked by embodied conversational agents.” [Online]. Available: http://www.speech.kth.se/snack/

[29] Z. Kmetty and K. Bozsonyi, “Identifying Depression-Related Behavior on Facebook—An Experimental Study,” *Social Sciences*, vol. 11, no. 3, Mar. 2022, doi: 10.3390/socsci11030135.

[30] S. Chancellor, J. Pater, T. Clear, E. Gilbert, and M. de Choudhury, “Thyghgapp: Instagram content moderation and lexical variation in Pro-Eating disorder communities,” in *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, Feb. 2016, vol. 27, pp. 1201–1213. doi: 10.1145/2818048.2819963.

[31] M. A. Moreno, A. Ton, E. Selkie, and Y. Evans, “Secret Society 123: Understanding the Language of Self-Harm on Instagram,” *Journal of Adolescent Health*, vol. 58, no. 1, pp. 78–84, Jan. 2016, doi: 10.1016/j.jadohealth.2015.09.015.

[32] R. C. Brown, T. Fischer, A. D. Goldwich, F. Keller, R. Young, and P. L. Plener, “#cutting: Non-suicidal self-injury (NSSI) on Instagram,” *Psychological Medicine*, vol. 48, no. 2, pp. 337–346, Jan. 2018, doi: 10.1017/S0033291717001751.

[33] F. Tang, I. Uchendu, F. Wang, H. H. Dodge, and J. Zhou, “Scalable diagnostic screening of mild cognitive impairment using AI dialogue agent,” *Scientific Reports*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-61994-0.

[34] R. Arrabales, “Perla: A Conversational Agent for Depression Screen-ing in Digital Ecosystems. Design, Implementation and Validation.”

[35] D. E. Losada and F. Crestani, “A Test Collection for Research on Depression and Language Use.” [Online]. Available: http://www.who.int/mediacentre/factsheets/fs369/en/

[36] J. J. Stephen and P. Prabu, “Detecting the magnitude of depression in Twitter users using sentiment analysis,” *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4, pp. 3247–3255, Aug. 2019, doi: 10.11591/ijece.v9i4.pp3247-3255.

[37] B. Keles, N. McCrae, and A. Grealish, “A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents,” *International Journal of Adolescence and Youth*, vol. 25, no. 1. Routledge, pp. 79–93, Jan. 02, 2020. doi: 10.1080/02673843.2019.1590851.

[38] D. William and D. Suhartono, “Text-based Depression Detection on Social Media Posts: A Systematic Literature Review,” in *Procedia Computer Science*, 2021, vol. 179, pp. 582–589. doi: 10.1016/j.procs.2021.01.043.

[39] A. Vij and J. Pruthi, “An automated Psychometric Analyzer based on Sentiment Analysis and Emotion Recognition for healthcare,” in *Procedia Computer Science*, 2018, vol. 132, pp. 1184–1191. doi: 10.1016/j.procs.2018.05.033.

[40] S. Almouzini, M. Khemakhem, and A. Alageel, “Detecting Arabic Depressed Users from Twitter Data,” in *Procedia Computer Science*, 2019, vol. 163, pp. 257–265. doi: 10.1016/j.procs.2019.12.107.

[41] E. Gaussier, IEEE Computational Intelligence Society, Association for Computing Machinery. Special Interest Group on Knowledge Discovery & Data Mining, Institute of Electrical and Electronics Engineers, and Association for Computing Machinery, *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics: IEEE/ACM DSAA’2015: 19-21 Oct 2015, Paris, France*.

[42] T. Dillon, Ed., *Artificial Intelligence in Theory and Practice IV*, vol. 465. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-25261-2.

[43] P. Mathur, R. Sawhney, S. Chopra, M. Leekha, and R. Ratn Shah, “Utilizing temporal psycholinguistic cues for suicidal intent estimation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12036 LNCS, pp. 265–271. doi: 10.1007/978-3-030-45442-5_33.

[44] S. Baker *et al.*, "Combatting social isolation and increasing social participation of older adults through the use of technology: A systematic review of existing evidence," *Australasian Journal on Ageing*, vol. 37, no. 3, pp. 184–193, Sep. 01, 2018. doi: 10.1111/ajag.12572.

[45] A. Amanat et al., "Deep Learning for Depression Detection from Textual Data," *Electronics*, vol 11, ID: 676, 2022, doi: 10.3390/ electronics11050676

[46] S. Jaiswal, M. Valstar, K. Kusumam, C. Greenhalgh, "Virtual Human Questionnaire for Analysis of Depression, Anxiety and Personality*," Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents,* 2019, doi:10.1145/3308532.3329469

[47] P. Philip, J.-A Micoulaud-Franchi, P. Sagaspe, "Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders," *Scientific Reports*, vol. 7, no. 1, 2017, doi:10.1038/srep42656

[48] A. Mallol-Ragolta, Z. Zhao, L. Stappen, N. Cummins, B. Schuller, "A hierarchical attention network-based approach for depression detection from transcribed clinical interviews," *Proc. Interspeech 2019*, 2019, pp. 221–225.

[49] B. Mirheidari, D. Blackburn, R. O'Malley, T. Walker, A. Venneri, M. Reuber, H. Christensen, "Computational cognitive assessment: Investigating the use of an intelligent virtual agent for the detection of early signs of dementia," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019, pp. 2732-2736, doi: 10.1109/ICASSP.2019.8682423..

[50] A. Khodabakhsh, S. Kuscuoglu, C. Demiroglu, "Detection of Alzheimer's disease using prosodic cues in conversational speech," in *2014 22nd Signal Processing and Communications Applications Conference (SIU)*, 2014, pp. 1003–1006, doi: 10.1109/SIU.2014.6830401.

[51] S. Jaiswal, M. Valstar, K. Kusumam, C. Greenhalgh, "Virtual human questionnaire for the analysis of depression, anxiety and personality," in *IVA 2019 - Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, Association for Computing Machinery, Inc, 2019, pp. 81–87, doi: 10.1145/ 3308532.3329469. 46

[52] Y.-T. Tsai, W.-A. Lin, "Design of an intelligent cognition assistant for people with cognitive impairment," in *Proceedings - 20th International Conference on High-Performance Computing and Communications, 16th International Conference on Smart City and 4th International Conference on Data Science and Systems*, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 1207–1212, doi: 10.1109/HPCC/SmartCity/DSS.2018.00203.

[53] H. Tanaka et al., "Detecting Dementia Through Interactive Computer Avatars," in IEEE Journal of *Translational Engineering in Health and Medicine*, vol. 5, pp. 1-11, 2017, Art no. 2200111, doi: 10.1109/JTEHM.2017.2752152.

[54] F. Tang, I. Uchendu, F. Wang, H. Dodge, J. Zhou, "Scalable diagnostic screening of mild cognitive impairment using ai dialogue agent," *Sci. Rep.*, vol. 10, 2020, doi: 10.1038/s41598-020-61994-0. 33

[55] R. Arrabales, Perla: A conversational agent for depression screening in digital ecosystems. design, implementation, and validation, *PsyArxiV Preprints*, pp. 1-16, 2020, doi: 10.31234/osf.io/nfas4.34

[56] D. Griol, J.M. Molina, Z. Callejas, "Combining speech-based and linguistic classifiers to recognize emotion in user spoken utterances," *Neurocomputing*, pp. 326–327, 2019, doi: 10.1016/j.neucom.2017.01.120. 22

[57] M. Kobayashi, A. Kosugi, H. Takagi, M. Nemoto, K. Nemoto, T. Arai, Y. Yamada, "Effects of age-related cognitive decline on elderly user interactions with voice-based dialogue systems," in *Lecture Notes in Computer Science*, 11749 LNCS, Springer Verlag, 2019, pp. 53–74, doi: 10.1007/978-3-030-29390-1_4.

[58] S. Okada, K. Inoue, T. Imai, M. Noguchi, K. Kuwamura, "Dementia scale classification based on ubiquitous daily activity and interaction sensing," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2019, pp. 192–198.

[59] B. Inkster, S. Sarda, V. Subramanian, "An empathy-driven, conversational artificial," *JMIR Mhealth Uhealth*, vol. 6, no. 11, 2018, doi: 10.2196/12106.

[60] T. Ujiro, H. Tanaka, H. Adachi, H. Kazui, M. Ikeda, T. Kudo, S. Nakamura, "Detection of dementia from responses to atypical questions asked by embodied conversational agents," in *19TH Annual Conference of the International Speech Communication Association (Interspeech 2018)*, vol. 1-6, Speech Research for Emerging Markets in Multilingual Societies, in Interspeech, Int Speech Commun Assoc, 2018, pp. 1691–1695, doi: 10.21437/Interspeech.20181514. 28

[61] P. Philip, J.-A. Micoulaud Franchi, P. Sagaspe, E. de Sevin, J. Olive, S. Bioulac, A. Sauteraud, "Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders," *Sci. Rep.*, vol. 7, 2017, doi: 10.1038/srep42656.

[62] M. Faurholt-Jepsen, J. Busk, M. Frost, M. Vinberg, E.M. Christensen, O. Winther, J.E. Bardram, L.V. Kessing, "Voice analysis as an objective state marker in bipolar disorder," *Transl. Psychiatry*, vol. 6, 2016, doi: 10.1038/tp.2016.

[63] M. Sidorov, W. Minker, Emotion recognition and depression diagnosis by acoustic and visual features: A multimodal approach, in *AVEC 2014 - Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, 2014, pp. 81–86, doi: 10.1145/2661806.2661816.
[64] A. Palanca, P. Flaschner, A. Thommandram, M. Li and Y. Fossat, “Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey,” *Journal of Medical Internet Research*, vol. 21, no. 4, p. e12887, 2019
[65] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, G. Lucas, S. Marsella, F. Morbini, A. Nazarian, S. Scherer, G. Stratou, A. Suri, D. Traum, R. Wood, Y. Xu, A. Rizzo, L.-P. Morency, “Simsensei kiosk: A virtual human interviewer for healthcare decision support, in 13th International Conference on Autonomous Agents and Multiagent Systems,” *International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS)*, 2014, pp. 1061–1068.
[66] Y. Imai, K. Hasegawa, “The revised Hasegawa's dementia scale (HDSR-R)-evaluation of its usefulness as a screening test for dementia,” *Hong Kong Journal of Psychiatry*, vol. 4, no. 2, pp. 20-24, 1994.
[67] NN, *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Ed., A.P. Association 2013, doi: 10.1176/appi.books.9780890425596.
[68] M. Ganguli, D. Blacker, D. Blazer, I. Grant, D. Jeste, J. Paulsen, R. Petersen, P. Sachdev, “Classification of neurocognitive disorders in DSM-5: A work in progress,” *The American Journal of Geriatric Psychiatry*, vol. 19, pp. 205–210, 2011, doi: 10.1097/JGP.0b013e3182051ab4.
[69] R.J. Castellani, R.K. Rolston, M.A. Smith, “Alzheimer's disease,” *Disease-a-Month*, vol. 56, no. 9, pp. 484-546, 2010, doi: 10.1016/j.disamonth.2010.06.001.
[70] Abd-Alrazaq, AA, Alajlani, M, Abdallah Alalwan, A et al., “An overview of the features of chatbots in mental health: A scoping review,” *International Journal of Medical Informatics*, vol. 132, 2019, doi: 10.1016/j.ijmedinf.2019.103978
[71] K. Schmidtke, S. Pohlmann, B. Metternich, “The syndrome of functional memory disorder: Definition, etiology, and natural course,” *The American Journal of Geriatric Psychiatry*, vol. 16, no. 12, pp. 981–988, 2008, doi: 10.1097/JGP.0b013e318187ddf9.
[72] C. Harada, M. Natelson Love, K. Triebel, “Normal cognitive aging,” *Clinics in Geriatric Medicine*, vol. 29, pp. 737–752, 2013, doi: 10.1016/j.cger.2013.07.002.
[73] Z. Zhang, "Microsoft Kinect Sensor and Its Effect," in *IEEE MultiMedia*, vol. 19, no. 2, pp. 4-10, Feb. 2012, doi: 10.1109/MMUL.2012.24.
[74] T.M. Mitchell, *Machine Learning*, McGraw-Hill, New York, 1997.
[75] C.M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berlin, Heidelberg, 2006.
[76] Y. Le Cun, Y. Bengio, G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015, doi: 10.1038/nature14539.
[77] D.R. Eignor, “The standards for educational and psychological testing,” *American Educational Research Association*, 2013, doi: 10.1037/14047-013.
[78] L.J. Cronbach, “Coefficient alpha and the internal structure of tests,” *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951, doi: 10.1007/BF02310555
[79] P. Rudd, “In search of the gold standard for compliance measurement,” *Archives of Internal Medicine*, vol. 139, pp. 627–628, 1979, doi: 10.1001/archinte.1979.03630430009004.
[80] I. Razykov, R. Ziegelstein, M. Whooley, B. Thombs, “The PHQ-9 versus the PHQ8-is item 9 useful for assessing suicide risk in coronary artery disease patients? Data from the heart and soul study,” *Journal of Psychosomatic Research*, vol. 73, pp. 163–168, 2012, doi: 10.1016/j.jpsychores.2012.06.001.
[81] Y. Wang, C. Gorenstein, “Assessment of depression in medical patients: A systematic review of the utility of the beck depression inventory-II,” *Clinics (São Paulo*), vol. 68, pp. 1274–1287, 2013, doi: 10.6061/clinics/2013(09)15.
[82] J. Tariman, D. Berry, B. Halpenny, S. Wolpin, K. Schepp, “Validation and testing of the acceptability E-scale for web-based patient-reported outcomes in cancer care,” *Applied Nursing Research*, vol. 24, no. 1, pp. 53–58, 2009, doi: 10.1016/j.apnr.2009.04.003.
[83] J. Brooke, “SUS - A quick and dirty usability scale,” *Project: System Usability Scale*, 2006.
[84] J. Martinez-Miranda, “Embodied conversational agents for the detection and prevention of suicidal behavior: Current applications and open challenges,” *Journal of Medical Systems*, vol. 41, no. 9, 2017, doi: 10.1007/s10916-017-0784-6.
[85] S. Baker, J. Warburton, J. Waycott, F. Batchelor, T. Hoang, B. Dow, E. Ozanne, F. Vetere, “Combatting social isolation and increasing social participation of older adults through the use of technology: A systematic review of existing evidence,” *Australasian Journal on Ageing*, vol. 37, no. 3, pp. 184-193, 2018, doi: 10.1111/ajag.12572.44

[86] S. H. Jayady and H. Antong, "Theme Identification using Machine Learning Techniques," *Journal of Integrated and Advanced Engineering (JIAE)*, vol. 1, no. 2, pp. 123-134, 2021, doi: 10.51662/jiae.v1i2.24

[87] K.T. Chui, W. Alhalabi, S.S.H. Pang, P. de Pablos, R.W. Liu, M. Zhao, Disease diagnosis in smart healthcare: Innovation, technologies, and applications, *Sustainability*, vol. 9, no. 12, 2017, doi: 10.3390/su9122309.

[88] A.A. Abd-al Razzaq, M. Alajlani, A.A. Alalwan, B.M. Bewick, P. Gardner, M. Househ, "An overview of the features of chatbots in mental health: A scoping review," *International Journal of Medical Informatics*, vol. 132, 2019, doi: 10.1016/j.ijmedinf.2019.103978.

[89] S. Hoermann, K.L. McCabe, D.N. Milne, R.A. Calvo, "Application of synchronous text-based dialogue systems in mental health interventions: Systematic review," *Journal of Medical Internet Research*, vol. 19, no. 8, pp. 91–100, 2017, doi: 10.2196/jmir.7023

[90] A. Ashraf, A. Sophian, A. A. Shafie, T. S. Gunawan, N. N. Ismail, A. A. Bawono, "Detection of Road Cracks Using Convolutional Neural Networks and Threshold Segmentation" *Integrated and Advanced Engineering (JIAE)*, vol. 2, no. 2, pp. 123-134, 2021, doi: 10.51662/jiae.v2i2.82

[91] R.A. Calvo, D.N. Milne, M.S. Hussain, H. Christensen, "Natural language processing in mental health applications using non-clinical texts," *Natural Language Engineering*, vol. 23, no. 5, pp. 649–685, 2017, doi: 10.1017/S1351324916000383.

[92] A. Russo, G. D'Onofrio, A. Gangemi, F. Giuliani, M. Mongiovi, F. Ricciardi, F. Greco, F. Cavallo, P. Dario, D. Sancarlo, V. Presutti, A. Greco, "Dialogue systems and conversational agents for patients with dementia: The human robot interaction," *Rejuvenation Research*, vol. 22, no. 2, pp. 109–120, 2019, doi: 10.1089/rej.2018.2075.

[93] K. T. Chui, W. Alhalabi, S. S. H. Pang, P. O. de Pablos, R. W. Liu, and M. Zhao, "Disease diagnosis in smart healthcare: Innovation, technologies and applications," *Sustainability (Switzerland)*, vol. 9, no. 12, 2017, doi: 10.3390/su9122309.

[94] A. A. Abd-alrazaq, M. Alajlani, A. A. Alalwan, B. M. Bewick, P. Gardner, and M. Househ, "An overview of the features of chatbots in mental health: A scoping review," *International Journal of Medical Informatics*, vol. 132, 2019, doi: 10.1016/j.ijmedinf.2019.103978.

[95] R. A. Calvo, D. N. Milne, M. S. Hussain, and H. Christensen, "Natural language processing in mental health applications using non-clinical texts," *Natural Language Engineering*, vol. 23, no. 5, pp. 649–685, Sep. 2017, doi: 10.1017/S1351324916000383.