Wearable artificial intelligence for anxiety and depression: A scoping review

Alaa Abd-alrazaq, Rawan AlSaad, Sarah Aziz, Arfan Ahmed, Kerstin Denecke, Mowafa Househ, Faisal Farooq, Javaid Sheikh

Submitted to: Journal of Medical Internet Research
on: September 13, 2022

Table of Contents

Wearable artificial intelligence for anxiety and depression: A scoping review

Alaa Abd-alrazaq[1] PhD; Rawan AlSaad[2] PhD; Sarah Aziz[1] MSc; Arfan Ahmed[1] PhD; Kerstin Denecke[3] Dr rer med; Mowafa Househ[4] PhD; Faisal Farooq[5] PhD; Javaid Sheikh[1] PhD

[1]AI Center for Precision Health, Weill Cornell Medicine-Qatar Doha QA
[2]College of Computing & Information Technology, University of Doha for Science and Technology Doha QA
[3]Institute for Medical Informatics, Bern University of Applied Science Bern CH
[4]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation Doha QA
[5]Qatar Computing Research Institute, Hamad bin Khalifa University Doha QA

Corresponding Author:
Alaa Abd-alrazaq PhD
AI Center for Precision Health, Weill Cornell Medicine-Qatar
P.O. Box 24144, Education City? Al Luqta St, Ar-Rayyan
Doha
QA

Abstract

Background: Anxiety and depression are the most common mental disorders worldwide. Owing to the lack of psychiatrists around the world, the incorporation of AI and wearable devices (wearable artificial intelligence (AI)) have been exploited to provide mental health services.

Objective: The current review aimed to explore the features of wearable AI used for anxiety and depression to identify application areas and open research issues.

Methods: We searched 8 electronic databases (MEDLINE, PsycINFO, EMBASE, CINAHL, IEEE Xplore, ACM Digital Library, Scopus, and Google Scholar). Then, we checked studies that cited the included studies, and screened studies that were cited by the included studies. Study selection and data extraction were carried out by two reviewers independently. The extracted data were aggregated and summarized using the narrative synthesis.

Results: Of the 1203 citations identified, 69 studies were included in this review. About two thirds of the studies used wearable AI for depression while the remaining studies used it for anxiety. The most frequent application of wearable AI was diagnosing anxiety and depression while no studies used it for treatment purposes. The majority of studies targeted individuals between the ages of 18 and 65. The most common wearable devices used in the studies were Actiwatch AW4. The wrist-worn devices were most common in the studies. The most commonly used data for model development were physical activity data, sleep data, and heart rate data. The most frequently used dataset from open sources was Depresjon. The most commonly used algorithms were Random Forest (RF) and Support Vector Machine (SVM).

Conclusions: Wearable AI can offer great promise in providing mental health services related to anxiety and depression. Wearable AI can be used by individuals as a pre-screening assessment of anxiety and depression. Further reviews are needed to statistically synthesize studies' results related to the performance and effectiveness of wearable AI. Given its potential, tech companies should invest more in wearable AI for treatment purposes for anxiety and depression.

(JMIR Preprints 13/09/2022:42672)
DOI: https://doi.org/10.2196/preprints.42672

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Wearable artificial intelligence for anxiety and depression: A scoping review

Alaa Abd-alrazaq[a]*; Rawan AlSaad[b], Sarah Aziz[a], Arfan Ahmed[a], Kerstin Denecke[c], Mowafa Househ[d], Faisal Farooq[e], Javaid Sheikh[a]
[a] AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar
[b] College of Computing & Information Technology, University of Doha for Science and Technology, Doha, Qatar
[c] School of Engineering and Computer Science, Institute for Medical Informatics, Bern University of Applied Sciences, Biel, Switzerland
[d] Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar
[e] Qatar Computing Research Institute, Hamad bin Khalifa University, Doha, Qatar
* Corresponding author.
E-mail address: aaa4027@qatar-med.cornell.edu

Abstract

Background: Anxiety and depression are the most common mental disorders worldwide. Owing to the lack of psychiatrists around the world, the incorporation of AI and wearable devices (wearable artificial intelligence (AI)) have been exploited to provide mental health services.

Objective: The current review aimed to explore the features of wearable AI used for anxiety and depression to identify application areas and open research issues.

Methods: We searched 8 electronic databases (MEDLINE, PsycINFO, EMBASE, CINAHL, IEEE Xplore, ACM Digital Library, Scopus, and Google Scholar). Then, we checked studies that cited the included studies, and screened studies that were cited by the included studies. Study selection and data extraction were carried out by two reviewers independently. The extracted data were aggregated and summarized using the narrative synthesis.

Results: Of the 1203 citations identified, 69 studies were included in this review. About two thirds of the studies used wearable AI for depression while the remaining studies used it for anxiety. The most frequent application of wearable AI was diagnosing anxiety and depression while no studies used it for treatment purposes. The majority of studies targeted individuals between the ages of 18 and 65. The most common wearable devices used in the studies were Actiwatch AW4. The wrist-worn devices were most common in the studies. The most commonly used data for model development were physical activity data, sleep data, and heart rate data. The most frequently used dataset from open sources was Depresjon. The most commonly used algorithms were Random Forest (RF) and Support Vector Machine (SVM).

Conclusion: Wearable AI can offer great promise in providing mental health services related to anxiety and depression. Wearable AI can be used by individuals as a pre-screening assessment of anxiety and depression. Further reviews are needed to statistically synthesize studies' results related to the performance and effectiveness of wearable AI. Given its potential, tech companies should invest more in wearable AI for treatment purposes for anxiety and depression.

Keywords: Wearable artificial intelligence; Artificial intelligence; Wearable devices; Anxiety; Depression; Scoping review

INTRODUCTION

Background

Anxiety and depression are amongst the “common mental illnesses” with high global prevalence. As of 2020, it has been reported that 19 percent of people worldwide suffered with depression or anxiety that prevents them to do their regular daily activities as they usually would for two weeks or longer[1]. In addition to having a significant economic impact on society[2], anxiety and depression affect people in terms of lost years because of illness. The statistics are mind blowing, depression is the world’s leading cause of disability within the youth population[3-5]. At 18 years of age, a previous study observed that depressed adults had 28 more years of quality-adjusted life expectancy (QALE) than non-depressed adults, resulting in a 28.9-year QALE loss due to depression in the United States[6]. Depression is also a significant risk factor when it comes to suicide[7]. The abovementioned statistics combined with the fact we only have around 9 psychiatrists per 100,000 people in developed countries[8] and 0.1 for every 1,000,000 in low-income countries[9] the situation is challenging to say the least. Current approaches for the assessment of anxiety and depression disorders are primarily based on clinical observations of patients' mental states, clinical history, and self-report questionnaires, such as the General Anxiety Disorder-7 (GAD-7) for anxiety and Patient Health Questionnaire-9 (PHQ-9) for depression. These methods are subjective, time-consuming, and challenging to repeat. As a result, contemporary psychiatric assessments can be inaccurate and ineffective at assessing anxiety and depression symptoms in a reliable and personalized manner. Therefore, there is a significant need to develop automatic techniques to address the limitations of the current psychiatric approaches for assessing anxiety and depression disorders and to overcome the shortages and uneven distribution of mental health professionals.

Recently, there have been rapid ongoing developments of artificial intelligence (AI) technology and wearables technology for healthcare and clinical use, offering numerous advantages towards individualizing diagnoses and treatment management of psychiatric disorders, including anxiety and depression [10-12]. Wearable technology includes electronic devices which users can wear near-body (e.g., smart watch, smart glasses, smart bracelet), on-body (e.g., electrocardiogram electrodes), in-body (e.g., implantable smart patch), and electronic textiles (e.g., smart clothes). Wearable devices are designed to provide a constant stream of health care data for disease diagnosis and treatment. This is achieved by continuously recording physiological parameters such as temperature, blood pressure, blood oxygen, respiratory rate, physical movement, and the electrical activity of the heart, brain, and skin. Symptoms of anxiety and depression can be assessed by many parameters collected in real-time by wearable devices for the diagnosis and monitoring of patients with anxiety and depression.

However, the dramatically accelerating pace in the development and adoption of wearables coupled with a shortage of skilled caregivers has led to an evolving need for automatic, efficient, and real-time approaches to analyze the large volumes of data collected by wearable sensors. This has motivated the integration of AI methods into wearable devices, introducing the “Wearable AI” technology. Wearable AI refers to intelligent electronic devices which are designed to be worn on the user’s body with intelligent operations. Wearable devices typically deal with monitoring and analyzing patients’ health data. However, when paired with AI, wearable AI introduces fundamental developments in the diagnosis and treatment of anxiety and depression. It has the potential to provide an early and accurate diagnosis of

anxiety and depression, facilitate more individualized treatment for anxiety and depression patients, and assist in developing preventative measures for groups at risk of anxiety and depression.

Research Problem and Aim

An extensive number of studies have been published on wearable devices combined with AI for anxiety and depression. Several reviews were conducted to summarize previous studies; however, they had the following limitations. Firstly, they focused on wearable devices rather than wearable devices paired with AI[10-15]. Secondly, they did not describe in detail the features of the used wearable devices and AI models[10-15]. Thirdly, they only targeted certain age groups such as children and adolescents[10,12]. Fourthly, they focused on wearable devices for either anxiety[11,14] or depression[12,13,15] rather than both anxiety and depression. Fifthly, they did not search relevant databases such as Medline[14], PsychInfo[10,13,15], IEEE Xplore[10-14], ACM Digital Library[10-15]. Lastly, they focused on wearables devices used for only diagnosing purposes using only ECG data[11] or EEG data[15]. Therefore, the need for a review that focuses on AI-paired wearable devices for anxiety and depression has never been higher. The review should be the same high-quality of a previous review conducted about AI-paired wearable devices for diabetes [16]. The current review aimed at exploring the features of wearable AI used for anxiety and depression, both to help customers make educated selections and to help the research community advance in this field by identifying gaps and looking into future prospects.

METHODS

To achieve the objective of the study, we carried out a scoping review consistent with Preferred Reporting Items for Systematic Reviews and Meta-Analyses- Extension for Scoping Reviews (PRISMA-ScR)[17]. PRISMA-ScR Checklist for this review is presented in Multimedia Appendix 1. The methods used in this review are detailed in the following subsections.

Search strategy

To find relevant studies, we searched 8 electronic databases on May 30, 2022: MEDLINE (via Ovid), PsycInfo (via Ovid), EMBASE (via Ovid), CINAHL (via EBSCO), IEEE Xplore, ACM Digital Library, Scopus, and Google Scholar. We set an automatic search biweekly for 24 weeks (ending on September 30, 2022). Given that Google Scholar retrieved a massive number of hits and order them based on their relevancy, only the first 100 hits (i.e.,10 pages) were checked in this review. To identify additional studies, we checked the reference lists of included studies (i.e., backward reference list checking) and screened studies that cited the included studies (i.e., forward reference list checking).

To develop the search query, three experts in digital mental health were consulted and previous reviews of relevance to the review were checked. The search query was composed of 3 groups of terms: terms related to AI (e.g., artificial intelligence, machine learning, and deep learning), terms related to wearable devices (e.g., wearable OR smart watch OR smartwatch), and terms related to anxiety and depression (e.g., anxiety OR anxious OR depression). Multimedia Appendix 2 presents the detailed search query used for searching each database.

Study Eligibility Criteria

This review included studies that focused on developing AI algorithms for anxiety and depression using data collected by wearable devices. Specifically, we focused on all AI algorithms used for any purpose related to anxiety and depression (e.g., diagnosis, monitoring, screening, therapy, predication, and prevention). The wearable devices that were used for collecting data had to be non-invasive on-body wearables such as smartwatches, smart glasses, smart clothing, smart bracelets, and smart tattoos. On the other hand, we excluded studies that used data collected by the following devices: non-wearable devices, hand-held devices (e.g., mobile phones), near-body wearable devices, in-body wearable devices (e.g., implants), wearable devices connected with non-wearable devices using wires, and wearable devices that need an expert to apply on users (e.g., wearable devices composed of many electrodes that need to be placed in very specific points of the body). Studies that used data collected via any methods (e.g., non-wearable devices, questionnaires, and interviews) in addition to wearable devices were considered in this review. We excluded studies that showed only a theoretical framework of AI-based wearable devices for anxiety and depression. We included journal articles, conference papers, and dissertations that were published in the English language since 2015. We excluded reviews, preprints, conference abstracts, posters, protocols, editorials, and commentaries. No restrictions were applied regarding the measured outcomes, setting, and country of publications.

Study Selection

We followed three steps in the study selection process. In the first step, we used EndNote X9 to remove duplicates from all retrieved studies. In the second step, we checked the titles and abstracts of the remaining publications. Lastly, we screened the entire texts of the studies included in the previous step. Two reviewers independently performed the study selection process. Disagreements between them in the second and third steps were resolved by discussion. Cohen's kappa was calculated to measure the inter-rater agreement [18], and it was 0.85 for "title and abstract" screening and 0.92 for full-text reading.

Data Extraction

Two reviewers utilized Microsoft Excel to independently extract data about study meta-data, wearable devices, and AI techniques. Any disagreements between the reviewers were resolved through discussion. The data extraction form used in this review was piloted using 5 studies, and it is shown in Multimedia Appendix 3.

Data Synthesis

Data that was extracted from the included studies were synthesized using the narrative approach, where data was summarized and described using texts, tables, and figures. To be more specific, we started by describing the meta-data of the included studies (e.g., year of publication and country of publication). Then, we presented the features of wearables devices used in the included studies (e.g., their status, type, placement, and operating system). Lastly, we summarized the characteristics of AI techniques used (e.g., AI algorithms used, their aim, dataset size, and data input type). We used Microsoft Excel to manage data synthesis.

RESULTS

Search Results

As depicted in Figure 1, searching all pre-identified databases retrieved 1203 records. Of these, 340 duplicates were detected and removed using reference management software (EndNote X9). Screening titles and abstracts of the remaining 863 citations resulted in excluding 506 records. We could find the full text of 7 records of the remaining 357 records. Reading the full text of the remaining 354 records led to excluding 298 records for several reasons shown in Figure 1. We identified 13 additional records relevant to this review by

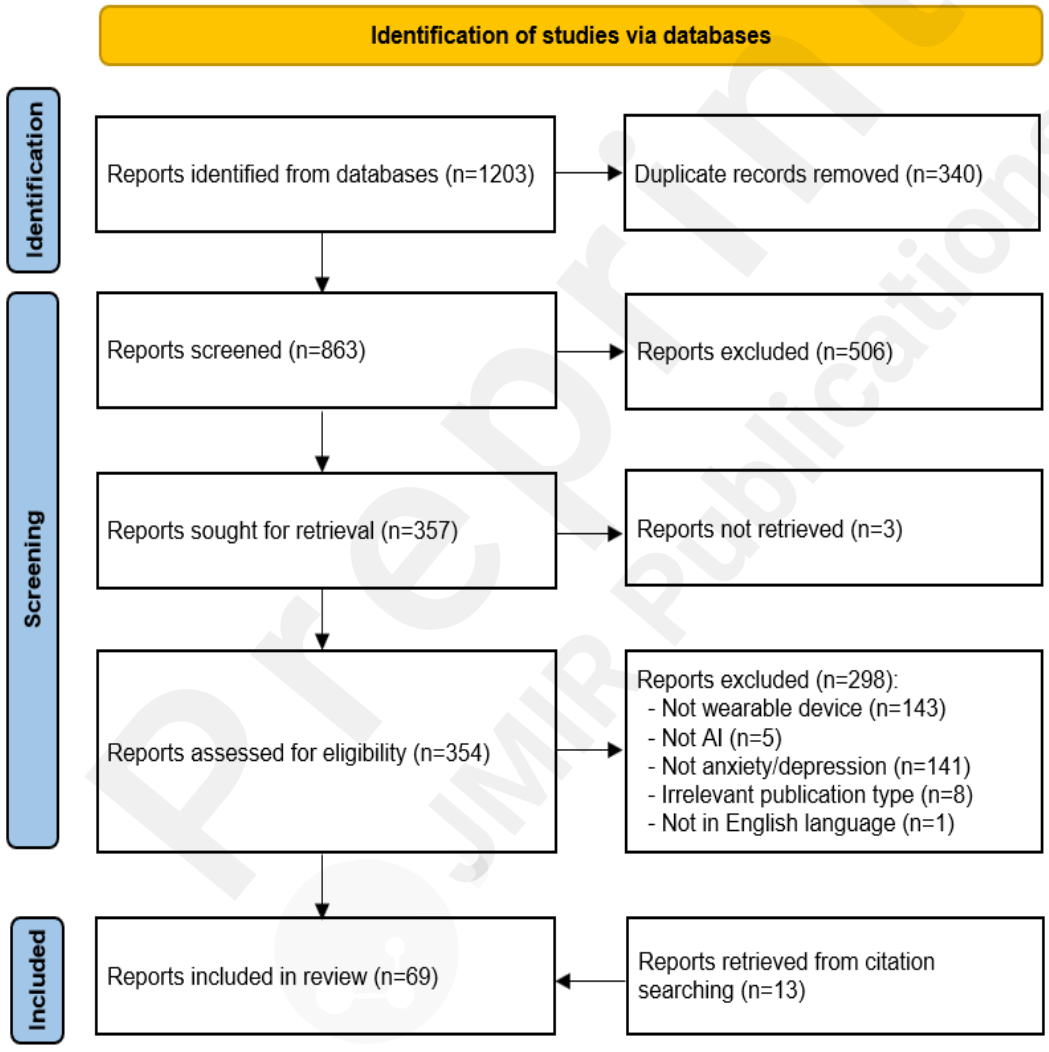


Figure 1: Flow chart of the study selection process

backward and forward reference list checking. In total, 69 records were included in the current review[19-87].

Characteristics of Included Studies

The included studies were published between 2015 and 2022 (Table 1). The years in which

the largest number of included studies was published were 2021 (17/69, 24.6%), 2019 (16/69, 23.2%), and then 2020 (15/69, 21.7%). Studies were carried out in 21 different countries (Table 1). More than a quarter of studies (21/69, 30.4%) were published in the United States. The included studies were peer-reviewed journal articles (49/69, 71%), conference proceedings (18/69, 26.1%), and theses (2/69, 2.9%).

Number of participants in the included studies ranged from 8 to 4036, with an average of 186.7 (standard deviation (SD)=522.2) (Table 1). The mean age of participants was reported in 50 studies and ranged between 5.2 and 78 years, with an average of 36.4 (SD 15.4). Only 6 of the included studies targeted children (<18 years), and 3 studies focused on only older adults (≥65 years). The percentage of female participants was reported in 54 studies and varied between 2.4% and 100%, with an average of 59.8 (SD 15.3). More than one-third of the studies (26/69, 37.7%) recruited participants with any health conditions, and about 30.4% of the studies (21/69) included both patients with depression and healthy individuals. Multimedia Appendix 4 shows characteristics of each included study.

Table 1: Characteristics of the included studies

Feature	Number of studies (%)	References
Year of publication		
2022	10 (13)	27,30,38,48,52,59,63,64,69,81
2021	17 (24.6)	19-21,23,25,28,41,45,49,54,61,62,68,73,74,77,78
2020	15 (21.7)	22,29,31,33,40,43,44,53,57,60,66,70,71,76,79
2019	16 (23.2)	26,32,34,42,46,47,51,56,65,67,72,75,80,84-86
2018	5 (7.2)	35,36,50,55,83
2017	4 (5.8)	24,37,39,58
2016	1 (1.4)	87
2015	1 (1.4)	82
Type of publication		
Journal article	49 (71)	19,21,23,25-30,34,38-46,48-54,56-61,64-66,69-71,73-75,77-79,81,82,84,86,87
Conference Paper	18 (26.1)	20,22,24,31-33,35-37,55,62,63,67,68,72,80,83,85
Thesis	2 (2.9)	47,76
Country of publication		
United States	21 (30.4)	24,25,30,31,37,41,42,50,54-56,59,61,66,74,76,77,80,83-85
Mexico	7 (10.1)	34,58,65,69-71,86
Norway	6 (8.7)	20,32,35,36,43,47
United Kingdom	5 (7.2)	29,38,48,72,78
South Korea	5 (7.2)	26-28,46,60
Japan	4 (5.8)	33,63,67,79
Pakistan	3 (4.3)	21,22,45
China	3 (4.3)	23,39,44
India	2 (2.9)	51,53
Taiwan	2 (2.9)	62,81
Others	11 (15.9)	19,40,49,52,57,64,68,73,75,82,87

Number of participants		
Mean (Standard Deviation)	186.9 (522.2)	19-87
Range	8-4036	19-87
1-100	53	19-22,24,26,27,30-40,42-49,53,55,56,58-68,70,71,75-79,81-83,85-87
101-500	11	23,25,41,50,51,54,57,69,73,80,84
>500	5	28,29,52,72,74
Age of participants		
Mean (Standard Deviation)	36.4 (15.44)	19-21,26-30,32,34-38,41-43,46-48,51,52,54-59,61-71,73-80,83,85,86
Range	5.2-78	19-21,26-30,32,34-38,41-43,46-48,51,52,54-59,61-71,73-80,83,85,86
<18	5 (7.5)	54-56,59,76
18-40	17 (24.6)	21,26,29,37,51,52,58,66,67,73-75,77,78,80,83,85
41-65	25 (36.2)	19,20,27,30,32,34-36,38,41-43,47,48,57,62-65,68-71,79,86
>65	3 (4.3)	27,46,61
Gender (Female %)		
Mean (Standard Deviation)	59.4 (15.64)	19-22,26-32,34-38,41-48,50,52,53,55-59,61-71,73,75-81,83,85,86
Range	2.4-100	19-22,26-32,34-38,41-48,50,52,53,55-59,61-71,73,75-81,83,85,86
Participant Health Conditions[1]		
Depression	32 (46.4)	19,20,23,24,26,30,32,34-38,42,43,46-48,50,57,59,60,62,65-71,77,79,86
Healthy	27 (39.1)	19,20,32,34-36,42,43,47,48,50,54-57,60,62,65,67-71,75,76,79,86
Any health condition	26 (37.7)	21,22,25,27-29,31,33,39,41,44,45,52,53,58,61,63,64,72-74,80,83-85,87
Internalizing disorders	4 (5.8)	54-56,76
Bipolar	3 (4.3)	26,49,82
Others	6 (8.7)	40,51,69,75,78,81

[1] number do not add up as participants in many studies have more than health condition

Features of Wearable Devices

The included studies focused on wearable devices for depression (44/69, 63.8%), anxiety (17/69, 24.6%), and both (8/69, 11.6%). Approximately 89.9% (62/69) of the included studies used commercial wearable devices (Table 2). The included studies used 41 different wearable devices. All studies used only one wearable device except 7 studies. The most common wearable devices used in the included studies were Actiwatch AW4 (17/69, 24.6%), Fitbit series (e.g., Fitbit Charge, Fitbit Flex, Fitbit Altra) (13/69, 18.8%), and Empatica series (e.g., E3 and E4) (7/69, 10.1%). The commercial wearable devices were manufactured by 25 different companies, but the most common companies were Cambridge Neurotechnology (17/69, 24.6%), Fitbit Inc (13/69, 18.8%), and Empatica (7/69, 10.1%). Multimedia Appendix 5 shows features of wearable devices in each included study.

The wearable devices in the included studies were available in 7 forms, but the most common forms were smart bands (50/69, 72.5%) and smartwatches (16/69, 23.2%) (Table 2). The wearable devices in the included studies were worn on 11 different parts of the body, but the

wrist-worn devices were most common in the included studies (57/69, 82.6%). The compatibility of the wearable devices with the operating systems of other devices was identified in 61 studies. The wearable devices were compatible with only one operating system in 41% (25/61) of studies and more than one operating system in 59% (36/61) of studies. The most common operating systems compatible with the wearables devices in the included studies were Windows (52/61, 85.2%) followed by iOS (36/61, 59%) and Android (35/61, 57.4%).

Only 21 studies (30.4%) used a gateway between the wearable device and the main host device (Table 2). In 13 of the 21 studies (61.9%), the gateway was PCs, smartphones, and tablets. The included studies used 4 types of host devices (i.e., end gate devices that stores data collected by the wearable devices). More than one host device was used in 14 studies (20.3%). The most common host devices in the included studies were computers (46/69, 66.7%) and database servers (30/69, 43.5%). Data is transferred from the wearable device to the host device through 6 different modes. In about 46.4% (32/69) of the studies, more than one mode of data transfer was used. The most common mode was Bluetooth (41/69, 59.4%) followed by docking stations (27/69, 39.1%) and Internet (24/69, 34.8%).

Table 2: Features of wearable devices

Feature	Number of studies (%)	References
Target Condition		
Depression	44 (63.8)	19,20,23-28,30,32,34-38,42,43,46-53,57,59,60,62,64-71,73,77,79,82-84,86
Anxiety	17 (24.6)	21,22,31,39-41,45,58,61,72,74,75,78,80,81,85,87
Anxiety and Depression	8 (11.6)	29,33,44,54-56,63,76
Status of WD[1]		
Commercial	63 (91.3)	19-38,40-43,46-60,62-71,73-81,83-87
Non-commercial	7 (10.1)	39,44,45,61,72,82,87
Name of WD[2]		
Actiwatch AW4	17 (24.6)	19,20,32,34-36,42,43,47,48,62,65,68-71,86
Fitbit series	13 (18.8)	25,26,30,31,33,38,50,52,59,63,73,80,84
Empatica series	7 (10.1)	27,37,58,66,75,78,85
3-Space Sensor	4 (5.8)	54-56,76
Muse	3 (4.3)	21,22,58
Others	29 (42.0)	23,24,28,29,39-41,44-46,49,51,53,57,58,60,61,64,67,72,74,77,79-83,85,87
Not reported	5 (7.2)	39,44,45,61,72
Company of WD[2]		
Cambridge Neurotechnology	17 (24.6)	19,20,32,34-36,42,43,47,48,62,65,68-71,86
Fitbit Inc	11(15.9)	25,26,30,31,33,38,50,52,59,63,73,80,84
Empatica	7 (10.1)	27,37,58,66,75,78,85
YEI Technology	4 (5.8)	54-56,76
InteraXon	3 (4.3)	21,22,58
Philips	3 (4.3)	41,46,57

Others	27 (39.1)	23,24,28,29,39,40,44,45,49,51,53,57,58,60,61,64,67,72,74,77,79-83,85,87
Not applicable	5 (7.2)	39,44,45,61,72,82
Type of WD[2]		
Smart band	50 (72.5)	21-26,28-31,33-40,42,44,45,47,50-59,61,63,66,69-76,78,79,81,84,85
Smartwatch	16 (23.2)	19,20,32,41,43,46,48,49,60,62,65,67,68,77,83,86
Others (smart shirt, smart adhesive electrodes, smart headset, smart glasses, smart ring, smart shirt)	5 (7.2)	64,80,82,85,87
Placement[2]		
Wrist	57 (82.6)	19,20,23-39,41-52,57-63,65-71,73-75,77-81,83-86
Head	7 (10.1)	21,22,53,54,58,76,87
Waist	6 (8.7)	28,54-56,72,76
Chest	4 (5.8)	58,80,82,85
Others (ankle, arm, eyes, finger, hand, neck, thigh)	1 (each) (1.4)	27,39,40,64,87
Compatibility with OS[3]		
Windows	52 (75.4)	19-22,25-27,30-38,40-43,46-50,52,53,57-59,61-63,65-75,78,80,81,83-87
IOS	36 (52.2)	21-31,33,37,38,50-53,58-60,63,64,66,73-75,77-81,83-85,87
Android	35 (50.7)	21-23,25-28,30,31,33,37,38,40,50-53,58-60,63,64,66,73-75,77-81,83-85,87
Mac OS	27 (39.1)	21,22,25-27,30,31,33,37,38,50,52,53,58,59,63,66,67,73-75,78,80,81,84,85,87
Linux	3 (4.3)	21,22,58
Not reported	8 (11.6)	39,44,45,54-56,76,82
Gateway[4]		
Smartphone	21 (30.4)	23,25,26,29-31,33,38,40,50,52,59-61,63,64,73,79,80,83,84
PC	13 (18.8)	25,26,30,31,33,38,50,52,59,63,73,80,84
Tablet	13 (18.8)	25,26,30,31,33,38,50,52,59,63,73,80,84
Silmee L20 gateway	1 (1.4)	79
Not reported	48 (69.6)	19-22,24,27,28,32,34-37,39,41-49,51,53-58,62,65-72,74-78,81,82,85-87
Host[5]		
PC	46 (66.7)	19-22,27,28,32,34-37,39-49,53-58,62,65-72,74-78,81,85-87
Server	30 (43.5)	23,25-27,29-31,33,37,38,50,52,58-61,63,64,66,73-75,78-85
Smartphone	16 (23.2)	21,22,24,27,37,51,53,58,66,74,75,77,78,81,85,87
Tablet	8 (11.6)	21,22,53,58,74,77,81,87
Mode of Data transfer[6]		
Bluetooth	41 (59.4)	21-27,29-31,33,37,38,40,50-56,58-61,63,64,66,73-81,83-85,87
Docking station	27 (39.1)	19,20,27,32,34-37,41-43,47-49,57,62,65-71,75,78,85,86
Internet	24 (34.8)	23,25,26,29-31,33,38,40,50,52,54-56,59,61,63,64,73,76,79,80,83,84
Removable media	8 (11.6)	39,44,45,54-56,76,82

Wired	8 (11.6)	28,46,54-56,58,72,76
ANT+	1 (1.4)	81

[1] number of studies does not add up as one study has one both commercial and non-commercial wearable devices.
[2] number of studies does not add up as several studies used more than one wearable device.
[3] number of studies does not add up as several studies used more than one wearable device, and many wearable devices are compatible with more than operating system (OS).
[4] number of studies does not add up as several studies used more than one wearable device, and many wearable devices used more than one gateway.
[5] number of studies does not add up as several studies used more than one wearable device, and many wearable devices used more than one host.
[6] number of studies does not add up as several studies used more than one wearable device, and many wearable devices used more than one of mode of data transfer.

Wearable devices measured more than one biosignal in 88.4% of the studies (61/69) (Table 3). The most commonly measured biosignals were physical activity measures (e.g., step counts, calories, distance, metabolic rate) (62/69, 89.9%), sleep measures (e.g., duration, patterns) (53/69, 76.8%), and heart rate measures (e.g., heart rate, heart rate variability, interbeat interval) (32/69, 46.4%). Wearable devices in the included studies contained 18 different sensors, and they contained more than one sensor in about 63.8% (44/69) of the included studies. The most common sensors in the wearable devices were accelerometers (63/69, 91.3%) and photoplethysmography (PPG) sensors (31/69, 44.9%). While wearable devices in 66.7% (46/69) of the studies used an opportunistic approach to collect data (i.e., automatic approach without user's input), they used both opportunistic approach and participatory approach (i.e., manual input by users) in the rest of the studies (23/69, 33.3%). Wearable devices in 55.1% (38/69) of the studies used a passive sensing method to collect data (i.e., the sensor captures only signals that come from an object without transmission of signals to it) whereas they used both passive sensing approach and active sensing approach (i.e., the sensor emits signals/light to an object, then captures the reflected signals/light via a detector to measure the biosignal) in the remaining studies (29/69, 44.9%). Multimedia Appendix 6 shows features of sensors of wearable devices in each included study.

Table 3: Features of sensors of wearable devices

Feature	Number of studies (%)	References
Measured biosignals[1]		
Physical activity measures	62 (89.9)	19,20,23-28,30-39,41-52,54-60,62-86
Sleep measures	53 (76.8)	19,20,23-27,30-38,41-43,46-52,57-60,62-71,73-75,77-86
Heart rate measures	32 (46.4)	23,26,27,29-31,33,37,38,40,50,51,58-61,63,64,66,72-75,77-83,85,87
Skin temperature	12 (17.4)	27,37,39,44,58,64,66,75,78,79,83,85
Electrodermal activity	11 (15.9)	27,37,40,58,61,66,72,75,78,83,85
Light exposure	7 (10.1)	28,41,46,49,57,77,83
Electroencephalograph	5 (7.2)	21,22,53,58,87
Respiratory	5 (7.2)	40,64,72,80,82
Audio	4 (5.8)	39,44,54,83
Electrocardiograph sensor	3 (4.3)	40,80,85
Ultraviolet level	3 (4.3)	64,79,83

Skin humidity	2 (2.9)	39,44
Air pressure	2 (2.9)	60,83
Others (blood oxygen saturation, location)	1 (each) (1.4)	40,81
Sensors in the wearables[2]		
Accelerometer	63 (91.3)	19,20,23-39,41-52,54-60,62-86
PPG sensors	31 (44.9)	23,26,27,29-31,33,37,38,40,50,51,58-61,63,64,66,72-75,77-81,83,85,87
Thermometer	12 (17.4)	27,37,39,44,58,64,66,75,78,79,83,85
Gyroscope	12 (17.4)	39,44,45,54-56,60,64,72,76,77,83
Electroencephalograph sensor	11 (15.9)	27,37,40,58,61,66,72,75,78,83,85
Altimeter	10 (14.5)	26,31,33,38,50,63,73,74,80,81
Light sensors	7 (10.1)	28,41,46,49,57,77,83
Electrocardiograph sensor	5 (7.2)	40,58,80,82,85
Compass	5 (7.2)	54-56,76,77
Microphone	4 (5.8)	39,44,54,83
Ultraviolet sensor	3 (4.3)	64,79,83
Barometer	2 (2.9)	60,83
Others (GPS, oximeter, piezoelectric sensor)	1 (each) (1.4)	40,81,83
Sensing approach[3]		
Opportunistic	69 (100)	19-87
Participatory	23 (33.3)	19,20,27,32,34-37,42,43,46-48,57,58,62,65,66,68-71,86
Sensing type[4]		
Passive	69 (100)	19-87
Active	31 (44.9)	23,26,27,29-31,33,37,38,40,50,51,58-61,63,64,66,72-75,77-81,83,85,87

[1] number of studies does not add up as several studies used more than one wearable device and most wearable devices assess more than one biosignal.
[2] number of studies does not add up as several studies used more than one wearable device and most wearable devices have more than one sensor.
[3] number of studies does not add up as several studies used more than one wearable device and many wearable devices used more than sensing approach.
[4] number of studies does not add up as several studies used more than one wearable device and many wearable devices used more than sensing type.

Features of AI Algorithms

The included studies used AI for 3 clinical purposes: (1) diagnosing or screening anxiety and depression (41/69, 59.4%), (2) monitoring symptoms or levels of anxiety and depression (15/69, 21.7%), and (3) predicting occurrence or level of anxiety and depression in the future based on previous and current biosignals (13/69, 18.8%) (Table 4). The included studies used

only machine learning algorithms (46/69, 66.7%), only deep learning algorithms (7/69, 10.1%), and both machine learning and deep learning algorithms (16/69, 23.2%). Studies used algorithms to solve classification problems (63/69, 91.3%), regression problems (11/69, 15.9%), and clustering problems (3/69, 4.3%). More than 50 different algorithms were used in the included studies, but the most commonly used algorithms were Random Forest (RF) (36/69, 52.2%), Support Vector Machine (SVM) (26/69, 37.7%), Logistic Regression (LogR) (16/69, 23.2%), Decision Tree (DT) (16/69, 21.7%), Extreme Gradient Boosting (XGBoost) (11/69, 15.9%), K-Nearest Neighbours (KNN) (11/69, 15.9%). Multimedia Appendix 7 shows features of AI algorithems in each included study.

Table 4: Features of AI algorithms

Feature	Number of studies (%)	References
AI category		
Machine Learning (ML)	46 (66.7)	19,20,23,25,26,31,33,34,37,39,40,42,46,49-61,63-67,69-71,73,75-81,83,84,86,87
Deep Learning (DL)	7 (10.1)	24,29,32,44,47,62,82
ML and DL	16 (23.2)	21,22,27,28,30,35,36,38,41,43,45,48,68,72,74,85
Problem-solving approaches[1]		
Classification	63 (91.3)	19-36,38-58,60-65,67-73,75,76,78-82,84-87
Regression	11 (15.9)	37,42,50,59,66,73,74,77,79,83,85
Clustering	3 (4.3)	31,74,85
AI Algorithm[2]		
Random Forest	36 (52.2)	19-23,26,27,30,33-38,41,43,45,46,49,51,53,59-61,64-66,68-71,77-79,81,86
Support Vector Machine	26 (37.7)	19,20,23,27,30,31,35,38,40,41,49,53,55,56,58,60,61,64,67,72,75,77-80,87
Logistic Regression	16 (23.2)	19,21-23,25,28,30,38,46,49,51,55-57,61,64
Decision Tree	16 (23.2)	20,23,27,35,38,40,46,49,54-56,72,76,78,81
Extreme Gradient Boosting	11 (15.9)	20,27,28,41,42,59,64,73,74,79,81
K-Nearest Neighbors	11 (15.9)	23,27,35,38,40,41,55,56,64,78,87
AdaBoost	9 (13.0)	25,30,35,37,59,68,77,81,84
Multilayer Perceptron	8 (11.6)	21,22,24,27,28,72,74,82
Convolutional Neural Network	7 (10.1)	32,43-45,47,48,62
Gradient Boosting	5 (7.2)	25,27,45,59,77
Naive Bayes	5 (7.2)	23,35,38,40,53
Others	28 (40.6)	19,28-31,35-37,40,41,43-45,47,48,50-53,59,63,66,68,74,77,81,83,85
Aim of AI algorithm		
Diagnosis/screening	41 (59.4)	19,21,22,28,32,35,36,38-40,43,46-49,51,53-57,61-63,65,67-71,73-76,78-80,82,83,85,87
Monitoring	15 (21.7)	20,23,27,34,37,42,44,45,50,58,60,64,66,72,86
Prediction	13 (18.8)	24-26,29-31,33,41,52,59,77,81,84
Ground Truth Assessment[3]		
MADRS	17 (24.6)	19,20,32,34-36,42,43,47,48,62,65,68-71,86
PHQ-4, -8, and -9	13 (18.8)	23,24,27,28,30,38,52,53,59,60,73,77,83
STAI	8 (11.6)	21,22,29,31,39,44,61,74
DSM-IV and -5	6 (8.7)	26,50,55,56,60,82

BDI-II	4 (5.8)	25,44,60,84
Others	26 (37.7)	27,29,33,37,40,41,45,46,49-51,54,57,58,63,64,66,69,75,76,78,79,81,82,85,87
Not reported	3 (4.3)	67,72,80
Validation approach[4]		
K-fold cross-validation	33 (47.8)	21-24,27,30,32,34,35,37,38,40,41,45,47,51,52,60,62,63,66,68,69,73-75,78-83,87
Hold-out cross-validation	25 (36.2)	26,28,29,31,32,34,37,44-46,48,49,51,60-62,66,67,70,71,74,81,82,84,86
Leave-one-out cross-validation	20 (29.0)	20,25,32,33,36,37,42,43,45,50,53-56,58,59,75,76,84,85
Nested cross-validation	3 (4.3)	19,64,77
External validation	1 (1.4)	57
Time-series cross-validation	1 (1.4)	64
Repeated random subsampling	1 (1.4)	87
Not reported	3 (4.3)	39,65,72
Performance measures[5]		
Accuracy	50 (72.5)	20-29,31-33,35,36,38-40,42,43,46-49,51,53-56,60-64,67-71,73,75,76,78,79,81,82,84,86-88
Sensitivity	41 (59.4)	19,21-23,26-28,32-36,38,41-43,46,47,51-54,56-58,60,62,64,65,67-73,79-81,84,86
F1-score	30 (43.5)	19-22,25,27,28,32,33,35,36,38,44,46,47,50-52,60-64,67-69,72,80,81,84
Specificity	28 (40.6)	19,21,26,32,34-36,41-43,46,47,51-54,56,58,62,65,67,70,71,73,79-81,86
Precision	24 (34.8)	19,22,28,32,33,35,36,38,46,47,51,53,58,60,62,64,67,68,70-73,84,86
Area Under the Curve	22 (31.9)	19,26,28,30,34,40,41,46,51,54-57,62,64,65,67,69,70,73,81,86
Mean Absolute Error	9 (13.0)	21,22,48,59,66,73,77,79,83
Matthews correlation coefficient	9 (13.0)	35,36,43,47,62,68,69
Cohen's Kappa	7 (10.1)	21,22,40,42,52,68,73
Root Mean Square Error	6 (8.7)	21,22,37,59,66,73
Balanced Accuracy	6 (8.7)	19,41,52,67,80,86
Receiver Operating Characteristic	6 (8.7)	19,27,55,65,81,86
Correlation coefficient (r)	5 (7.2)	42,66,74,79,83
Others	13 (18.8)	22,40,50,52,53,57,59,71,73,74,77,85,86

[1] number of studies does not add up as many studies used more than one problem-solving approach.
[2] number of studies does not add up as many studies used more than one AI algorithm.
[3] number of studies does not add up as many studies used more than one tool to assess the ground truth.
[4] number of studies does not add up as many studies used more than one validation approach.
[5] number of studies does not add up as most studies used more than one performance measures.

The included studies identified the ground truth based on 27 different tools, but the most common tools were Montgomery-Asberg Depression Rating Scale (MADRS) (17/69, 24.6%), Patient Health Questionnaire-9 (PHQ-9) (12/69, 17.4%), and State-Trait Anxiety Inventory (STAI) (8/69, 11.6%). The included studies used 7 different validation methods of the models. About 21.7% (15/69) of the included studies used more than validation methods (Table 4). The most commonly used validation methods were K-fold cross-validation (33/69, 47.8%), hold-out cross-validation (25/69, 36.2%), and leave-one-out cross-validation (LOOCV) (20/69, 29%). The included studies evaluated the performance of the models using 33 different metrics. The most common metrics used in the included studies were accuracy (50/69, 72.5%), sensitivity (41/69, 59.4%), F1-score (30/69, 43.5%), specificity (28/69, 40.6%), precision (24/69, 34.8%), and area under the curve (AUC) (22/69, 31.9%).

About 20.3% (14/69) of the included studies reported the dataset size used for developing (i.e., training and testing) the models (Table 5). The dataset size ranged between 168 and 1570144 inputs, with an average of 168023 (SD=428843). The included studies used datasets from either closed sources (i.e., collected by authors of the study or obtained from previous studies) (50/69, 72.5%) or open sources (i.e., public databases) (19/69, 27.5%). Depresjon was the most common dataset obtained from open sources and used in the included studies (16/19, 84.2%). In 59.4% (41/69) of the studies, AI algorithms were developed using data collected by only wearable devices. Around 17.4% (12/69) of the studies developed AI algorithms using data collected by a combination of wearable devices and self-administered questionnaires (i.e., self-reported data). About 13% (9/69) of the studies developed AI algorithms using data collected by a combination of wearable devices and non-wearable devices (e.g., smartphones). Around 10.1% (7/69) of the studies developed AI algorithms using data collected by a combination of wearable devices, non-wearable devices, and self-administered questionnaires. The included studies used more than 50 categories of data to develop the model. While 43.5% (30/69) of the studies used only one category of the data to develop their models, the rest of the studies (39/69, 56.5%) used more than one category of the data. The most common data used to develop the models were physical activity data (e.g., step counts, calories, metabolic rate) (53/69, 76.8%), sleep data (e.g., duration, patterns) (27/69, 39.1%), heart rate data (e.g., heart rate, heart rate variability, interbeat interval) (26/69, 37.7%), mental health measures (e.g., depression level, anxiety level, stress level, mood status) (14/69, 20.3%), location data (e.g., latitude, longitude, % of time at home, stationary time) (10/69, %14.5), smartphone usage data (e.g., display on/off, charging activity, number of apps used) (10/69, %14.5), and social interaction (e.g., call and message logs) (10/69, %14.5). Number of features used in the model development ranged from 2 to 5173. In about half of studies (33/69, 47.8%), number of features was 10 or lower. Multimedia Appendix 8 shows features of data used for AI development in each included study.

Table 5: Features of data used for AI development

Feature	Number of studies (%)	References
Dataset size		
Mean (Standard Deviation)	168022.5 (428843.2)	22,23,28,37,41,44,45,51,58,60-62,70,73
Range	168-1570144	22,23,28,37,41,44,45,51,58,60-62,70,73
Dataset source		
Open	19 (27.5)	19,20,28,31,32,34,36,42,43,47,48,62,65,68-71,74,86
Closed	50 (72.5)	21-27,29,30,33,35,37-41,44-46,49-61,63,64,66,67,72,73,75-85,87
Data types		
Wearable device (WD)-based	41 (59.4)	20-22,27,29,31-36,38,39,41-48,53-56,58,61,62,65,67,69-71,73,75,76,78-80,82,87
WD-based, self-reported	12 (17.4)	19,26,28,30,49,51,52,57,68,81,85,86
WD-based, non-WD based	9 (13.0)	23,25,40,50,59,66,72,74,84
WD-based, non-WD based, self-reported	7 (10.1)	24,37,60,63,64,77,83
Data input to AI algorithm[1]		
Physical activity data	53 (76.8)	19,20,23-27,30-32,34-38,41-51,54-57,59,60,62-74,76,77,79,81,83-86
Sleep data	27 (39.1)	23-26,30,33,37,38,41,46,49-52,57,59,60,63,64,66,73,74,77,79,81,83,84

Heart rate data	26 (37.7)	23,26,27,29-31,40,50,51,58-61,63,64,66,72,75,77-81,83,85,87
Mental health measures	14 (20.3)	24,26,30,37,46,49,50,52,57,60,64,77,81,85
Social interaction data	10 (14.5)	23-25,37,59,60,66,72,83,84
Location data	10 (14.5)	23,25,37,50,59,64,66,74,83,84
Smartphone usage data	10 (14.5)	23,25,37,59,60,64,66,74,83,84
Electrodermal activity data	10 (14.5)	27,37,40,58,61,66,72,75,78,85
Skin temperature data	5 (7.2)	27,75,78,79,85
Demographic data	5 (7.2)	30,52,57,68,85
Electroencephalograph data	4 (5.8)	21,22,53,87
Light exposure	4 (5.8)	26,46,60,79
Audio data	4 (5.8)	39,44,54,85
Others	17 (24.6)	24,28,30,37,49,52,57,60,63,66,72-74,77,81,82,85
Number of features[2]		
1-10	33 (47.8)	19,21-25,27,34-40,43,46,47,50,54-58,67,69-72,75,78,82,83,87
11-20	16 (23.2)	23,26,28,30,33,45,48,51-53,57,61,68,72,76,86
21-30	6 (8.7)	44,52,60,63,73,85
31-40	6 (8.7)	23,34,38,50,66,73
41-50	6 (8.7)	23,41,64,73,77,80
>50	8 (11.6)	23,27,59,73,74,79,81,84
Not reported	8 (11.6)	20,29,31,32,42,49,62,65

[1] number of studies does not add up as many studies used more than one data input.
[2] number of studies does not add up as several studies used various numbers of features.

DISCUSSION

Principle Findings

This scoping review aimed at exploring features of AI and wearable devices used for anxiety and depression. In this review, about two thirds of the studies used wearable AI for depression while the remaining studies used it for anxiety. This may be attributed to the capabilities of wearables to collect biosingals related to symptoms of depression and anxiety. More specifically, it is well known that depression is associated with a decrease in activity and changes in sleep behaviours[13,89,90], which can be objectively measured by wearable devices. Further, analysis of depression symptoms does not rely upon highly accurate data; that is, general trends are sufficient to provide indications. In contrast, anxiety is usually associated with heart rate variability[91]. Although wearable devices can have an acceptable heart rate accuracy[92], the quality differs among devices[93]. Beyond, monitoring the heart rate without context information might be misleading since multiple factors impact the heart rate, thus, detecting anxiety based on only objective biosingals is questionable. Combination with additional data sources is crucial. So far, only a few studies in this review are based upon a combination of data from different sources (i.e., wearable devices, non-wearable devices, and self-administered questionnaires).

In this review, the most frequent application of wearable AI is diagnosing or screening anxiety and depression. A similar result was reported by 2 previous reviews, which showed that most studies focused on using wearables for diagnostic purposes[10,13]. Although wearable AI can be used for interventional and treatment purposes (e.g., personalized mindfulness,

meditation, and biofeedback therapy[14]), none of the systems in included studies was used for such purposes. This may be attributed to the lack of evidence on the effectiveness of wearable AI for improving anxiety and depression.

Smart bands worn on the wrist were most often applied in the studies. This has already been indicated by previous reviews as well[10,13,14]. This can may be attributed to the fact that wrist-worn wearable devices are less distractive and obtrusive, easy to use, and more stylish and familiar to most people. According to Hunkin et al.[94], such features are crucial for users' acceptance and use of wearable devices.

The most commonly used data for model development were physical activity data, sleep data, and heart rate data. This is expected given that depression and anxiety are associated with physical activity[13,89,90], sleep pattens[13,95,96], and heart rate[91], in addition, as the current review showed, these are the most common biosignals measured by commercial wearable devices.

Surprisingly, more than half of the papers considered only data from wearables in their AI algorithms. However, wearables cannot detect all symptoms of relevance for anxiety and depression for 2 reasons. Firstly, wearable devices cannot detect several physiological data such as weight loss or gain and changes in appetite[13]. Secondly, wearable devices cannot evaluate subjective symptoms such as social interaction, medical history, and lifestyle changes[13]. We might question whether research starts to place overreliance upon the diagnostic and predictive power of data from wearable devices only.

About one-fourth of studies relied upon a dataset called Depresjon[35] to develop their models. Depresjon is a freely available dataset that contains data related to the motor activity measured using an actigraph watch worn at the wrist (Actiwatch AW4, Cambridge Neurotechnology Ltd)[35]. The dataset also contains data related to depression levels assessed using the MADRS[35]. This explains why the most common wearable device used in the included studies was Actiwatch AW4 and why MADRS was the most frequently used tool to assess the ground truth.

Regarding the target population, we have to recognize that the majority of studies addressed individuals between the ages of 18 and 65. Global statistics show that depression and anxiety occur all over the age ranges starting at 15 with almost the same percentage. Only for adults at an age of 65 and older, there is a decrease in the percentage[1]. This might explain why the studies mainly targeted the age group 18 to 65. Another explanation might be that wearables are more popular for adults in that age range.

This review showed that K-fold cross-validation was the most frequently used validation method. This can be attributed to several reasons. Firstly, in comparison with hold-out cross-validation, K-fold cross-validation is prone to less variation as each observation is used for both training and testing. Secondly, the training set in K-fold cross-validation is larger than the training set in hold-out cross-validation, thereby, K-fold cross-validation has reduced bias and reduced over-estimation of test-error. Lastly, K-fold cross-validation is less expensive computationally than LOOCV as the algorithm needs to rerun only k times (usually ≤10).

Research and Practical Implications

The performance of wearable AI in diagnosing, monitoring, and predicting anxiety and depression was not assessed in this review. Systematic reviews and meta-analyses are needed to examine its performance. Future studies should also compare the performance of different wearable devices (e.g., Fitbit vs. Empatica), worn at different placements (e.g., wrist, chest, waist), and using different data types (e.g., wearable based-data vs. wearable based-data and self-reported data). Conducting systematic reviews of such studies can help researchers, developers, and wearable device companies to identify the most significant features and

powerful AI algorithms in diagnosing, monitoring, and predicting anxiety and depression.

AI research highly depends on available datasets. However, when only one dataset is exploited by researchers, no conclusions regarding the generalizability of study results can be drawn. Therefore, we recommend researchers (1) publish their datasets in open databases after ensuring participants' privacy and confidentiality and (2) exploit different datasets available in open databases.

The current review found a lack of AI-based wearable devices used for treatment purposes although wearable AI can be used for providing many interventions for anxiety and depression such as personalized mindfulness, meditation, and biofeedback therapy. Tech companies should invest more in wearable AI for treatment purposes for anxiety and depression. Researchers should also assess the effectiveness of such technologies in improving anxiety and depression.

The ground truth of mental states (anxiety or depression) in included studies was identified based on 27 different tools. Although most of these tools have been validated extensively, they usually do not include physiological biomarkers (e.g., physical activities, heart rate, EDA, respiratory rate, EEG). This brings into question the validity and reliability of drawing conclusions about mental states (anxiety or depression) based on physiological biomarkers when the grand truth of mental states is assessed using subjective questionnaires. Accordingly, the performance of AI-based wearable devices will be underestimated.

Although the current studies showed that wearable AI can be used for monitoring symptoms or levels of anxiety and depression, continuous tracking of physiological biomarkers could trigger emotional instability and ruminative thinking[97]. Although the wearable AI can approximate mental states (e.g., feeling nervous, anxious, or on edge) through heart rate and other variables, it could provide many false positives, thereby, exacerbating or increasing the anxiety or depression of an individual. The above-mentioned downsides of wearable AI should be considered and mitigated before developing AI-based wearables. More research studies are needed on the use of wearable devices and their impact on individual emotional and behavioural responses to a wearable device's automated feedback.

Wearable AI can help individuals conduct mental health and well-being pre-screening assessments without an initial hospital or clinical encounter. The individual could be notified through the wearable device, smartphone, or desktop application about their mental health status which would encourage them to visit a mental health and well-being professional. Such pre-screening feedback from wearables may help reduce mental health stigma and allow a higher number of individuals to seek help from a mental health professional.

The quality of the data, whether it is obtained from open sources or generated from wearable devices, should be emphasized. To do so, there is a need to be more practical standards for wearable device development that ensures accurate measurement of different signals generated from wearable devices to improve algorithmic performance.

Limitations

This review excluded many studies that focused on non-wearable devices, hand-held devices (e.g., mobile phones), near-body wearable devices, in-body wearable devices (e.g., implants), wearable devices connected with non-wearable devices using wires, and wearable devices that need an expert to apply on users. For this reason, our findings may not be generalizable to contexts where such excluded devices are applied. Owing to practical constraints, we included only studies published in the English language. We also restricted our search to studies published from 2015 onwards given that this is a fast-growing field, thereby, studies published before 2015 can be deemed outdated. Consequently, it is likely that we missed

some studies published in other languages and/or published before 2015. Another limitation of this review is that we cannot comment on the performance of wearable AI in diagnosing, monitoring, and predicting anxiety and depression and the importance of features/variables as this is out of the scope of the current review and needs systematic reviews, where the quality of the evidence and risk of bias are assessed.

CONCLUSION

Wearable AI can offer great promise in providing mental health services related to anxiety and depression. Wearable AI can be used by individuals as a pre-screening assessment of anxiety and depression. Further reviews are needed to statistically synthesize studies' results related to the performance and effectiveness of wearable AI. More studies are needed on the use of wearable devices and their impact on individual emotional and behavioural responses to a wearable device's automated feedback. Given its potential, tech companies should invest more in wearable AI for treatment purposes for anxiety and depression. Downsides of wearable AI (e.g., false positive alerts and triggering emotional instability and ruminative thinking) should be considered and mitigated before developing it.

ACKNOWLEDGEMENTS

No relevant funding

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Alaa Abd-alrazaq, Arfan Ahmed, and Sarah Aziz developed the protocol with guidance from and under the supervision of Javaid Sheikh. Alaa Abd-alrazaq searched the electronic databases and conducted backward and forward reference list checking. The study selection process was carried out by Alaa Abd-alrazaq & Rawan AlSaad. The data extraction process was conducted by Rawan AlSaad and Sarah Aziz. Alaa Abd-alrazaq and Sarah Aziz conducted data synthesis. Alaa Abd-alrazaq wrote results and methods sections. Dr. Arfan and Alaa Abd-alrazaq wrote the introduction section. Kerstin Denecke, Alaa Abd-alrazaq, Mowafa Househ, and Faisal Farooq wrote the discussion section. The article was revised critically for important intellectual content by all authors. All authors approved the manuscript for publication and agree to be accountable for all aspects of the work.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request

CODE AVAILABILITY

No custom code or mathematical algorithm was used in this study.

ADDITIONAL INFORMATION

Supplementary information: The online version contains supplementary material available at

REFERENCES

1 Wellcome Trust. Percentage of respondents worldwide who felt anxiety or depression for more than two weeks as of 2020, by country income and age. (2020).

2 Konnopka, A. & König, H. Economic Burden of Anxiety Disorders: A Systematic Review and Meta-Analysis. *Pharmacoeconomics* **38**, 25-37, doi:10.1007/s40273-019-00849-7 (2020).

3 Hammen, C., Brennan, P. A. & Keenan-Miller, D. Patterns of adolescent depression to age 20: the role of maternal depression and youth interpersonal dysfunction. *J Abnorm Child Psychol* **36**, 1189-1198, doi:10.1007/s10802-008-9241-9 (2008).

4 Whiteford, H. A. *et al.* Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575-1586, doi:10.1016/s0140-6736(13)61611-6 (2013).

5 Wilson, S. *et al.* Age of onset and course of major depressive disorder: associations with psychosocial functioning outcomes in adulthood. *Psychol Med* **45**, 505-514, doi:10.1017/s0033291714001640 (2015).

6 Jia, H. *et al.* Impact of depression on quality-adjusted life expectancy (QALE) directly as well as indirectly through suicide. *Soc Psychiatry Psychiatr Epidemiol* **50**, 939-949, doi:10.1007/s00127-015-1019-0 (2015).

7 Hawton, K., Casañas, I. C. C., Haw, C. & Saunders, K. Risk factors for suicide in individuals with depression: a systematic review. *J Affect Disord* **147**, 17-28, doi:10.1016/j.jad.2013.01.004 (2013).

8 Murray, C. J. *et al.* Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2197-2223, doi:10.1016/s0140-6736(12)61689-4 (2012).

9 Oladeji, B. D. & Gureje, O. Brain drain: a challenge to global mental health. *BJPsych Int* **13**, 61-63, doi:10.1192/s2056474000001240 (2016).

10 Welch, V. *et al.* Use of Mobile and Wearable Artificial Intelligence in Child and Adolescent Psychiatry: Scoping Review. *J Med Internet Res* **24**, e33560, doi:10.2196/33560 (2022).

11 Elgendi, M. & Menon, C. Assessing Anxiety Disorders Using Wearable Devices: Challenges and Future Directions. *Brain Sci* **9**, doi:10.3390/brainsci9030050 (2019).

12 Sequeira, L. *et al.* Mobile and wearable technology for monitoring depressive symptoms in children and adolescents: A scoping review. *J Affect Disord* **265**, 314-324, doi:10.1016/j.jad.2019.11.156 (2020).

13 Lee, S., Kim, H., Park, M. J. & Jeon, H. J. Current Advances in Wearable Devices and Their Sensors in Patients With Depression. *Front Psychiatry* **12**, 672347, doi:10.3389/fpsyt.2021.672347 (2021).

14 Juchems, P. *The use of wearable devices in the treatment and detection of anxiety: a systematic scoping review.* Master degree thesis, University of Twente, (2022).

15 Yasin, S. *et al.* EEG based Major Depressive disorder and Bipolar disorder detection using Neural Networks:A review. *Computer Methods and Programs in Biomedicine* **202**, 106007, doi:https://doi.org/10.1016/j.cmpb.2021.106007 (2021).

16 Ahmed, A. *et al.* Overview of Artificial Intelligence-Driven Wearable Devices for Diabetes: Scoping Review. *J Med Internet Res* **24**, e36010, doi:10.2196/36010 (2022).

17 Tricco, A. C. *et al.* PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine* **169**, 467-473, doi:10.7326/M18-0850 (2018).

18 Higgins, J. & Deeks, J. in *Cochrane Handbook for Systematic Reviews of Interventions* (eds J. Higgins & S. Green) (John Wiley & Sons, 2008).

19 Adamczyk, J. & Malawski, F. Comparison of Manual and Automated Feature Engineering for Daily Activity Classification in Mental Disorder Diagnosis. *Computing and Informatics* **40**, 850–879-850–879 (2021).

20 Aminifar, A., Rabbi, F., Pun, V. K. I. & Lamo, Y. Monitoring Motor Activity Data for Detecting Patients' Depression Using Data Augmentation and Privacy-Preserving Distributed Learning. *Annu Int Conf IEEE Eng Med Biol Soc* **2021**, 2163-2169, doi:10.1109/embc46164.2021.9630592 (2021).

21 Arsalan, A. & Majid, M. A study on multi-class anxiety detection using wearable EEG headband. *Journal of Ambient Intelligence and Humanized Computing*, doi:10.1007/s12652-021-03249-y (2021).

22 Arsalan, A., Majid, M. & Anwar, S. M. in *Intelligent Technologies and Applications.* (eds Imran Sarwar Bajwa, Tatjana Sibalija, & Dayang Norhayati Abang Jawawi) 187-197 (Springer Singapore).

23 Bai, R. *et al.* Tracking and Monitoring Mood Stability of Patients With Major Depressive Disorder by Machine Learning Models Using Passive Digital Data: Prospective Naturalistic Multicenter Study. *JMIR Mhealth Uhealth* **9**, e24365, doi:10.2196/24365 (2021).

24 Bennett, C. C. *et al.* in *2017 IEEE International Conference on Healthcare Informatics (ICHI)*. 536-540.

25 Chikersal, P. *et al.* Detecting Depression and Predicting its Onset Using Longitudinal Symptoms Captured by Passive Sensing: A Machine Learning Approach With Robust Feature Selection. *ACM Trans. Comput.-Hum. Interact.* **28**, Article 3, doi:10.1145/3422821 (2021).

26 Cho, C. H. *et al.* Mood Prediction of Patients With Mood Disorders by Machine Learning Using Passive Digital Phenotypes Based on the Circadian Rhythm: Prospective Observational Cohort Study. *J Med Internet Res* **21**, e11029, doi:10.2196/11029 (2019).

27 Choi, J. *et al.* Depressed Mood Prediction of Elderly People with a Wearable Band. *Sensors (Basel)* **22**, doi:10.3390/s22114174 (2022).

28 Choi, J. G., Ko, I. & Han, S. Depression Level Classification Using Machine Learning Classifiers Based on Actigraphy Data. *IEEE Access* **9**, 116622-116646, doi:10.1109/ACCESS.2021.3105393 (2021).

29 Coutts, L. V., Plans, D., Brown, A. W. & Collomosse, J. Deep learning with wearable based heart rate variability for prediction of mental and general health. *J Biomed Inform* **112**, 103610, doi:10.1016/j.jbi.2020.103610 (2020).

30 Dai, R. *et al.* Multi-Task Learning for Randomized Controlled Trials: A Case Study on Predicting Depression with Wearable Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **6**, Article 50, doi:10.1145/3534591 (2022).

31 Feng, T. & Narayanan, S. S. in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* 1011-1015.

32 Frogner, J. I. *et al.* in *Proceedings of the 4th International Workshop on Multimedia for Personal Health & Health Care* 9–15 (Association for Computing Machinery, Nice, France, 2019).

33 Fukuda, S. *et al.* in *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops).* 1-6.

34 Galván-Tejada, C. E. *et al.* Depression Episodes Detection in Unipolar and Bipolar Patients: A Methodology with Feature Extraction and Feature Selection with Genetic

Algorithms Using Activity Motion Signal as Information Source. *Mobile Information Systems* **2019**, 8269695, doi:10.1155/2019/8269695 (2019).

35 Garcia-Ceja, E. *et al.* in *Proceedings of the 9th ACM Multimedia Systems Conference* 472–477 (Association for Computing Machinery, Amsterdam, Netherlands, 2018).

36 Garcia-Ceja, E. *et al.* in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS).* 316-321.

37 Ghandeharioun, A. *et al.* in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII).* 325-332.

38 Griffiths, C. *et al.* Investigation of physical activity, sleep, and mental health recovery in treatment resistant depression (TRD) patients receiving repetitive transcranial magnetic stimulation (rTMS) treatment. *J Affect Disord Rep* **8**, 100337, doi:10.1016/j.jadr.2022.100337 (2022).

39 Gu, J. *et al.* Wearable Social Sensing: Content-Based Processing Methodology and Implementation. *IEEE Sensors Journal* **17**, 7167-7176, doi:10.1109/JSEN.2017.2754289 (2017).

40 Ihmig, F. R. *et al.* On-line anxiety level detection from biosignals: Machine learning based on a randomized controlled trial with spider-fearful individuals. *PLoS One* **15**, e0231517, doi:10.1371/journal.pone.0231517 (2020).

41 Jacobson, N. C., Lekkas, D., Huang, R. & Thomas, N. Deep learning paired with wearable passive sensing data predicts deterioration in anxiety disorder symptoms across 17-18 years. *J Affect Disord* **282**, 104-111, doi:10.1016/j.jad.2020.12.086 (2021).

42 Jacobson, N. C., Weingarden, H. & Wilhelm, S. Digital biomarkers of mood disorders and symptom change. *npj Digital Medicine* **2**, 3, doi:10.1038/s41746-019-0078-0 (2019).

43 Jakobsen, P. *et al.* Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls. *PLoS One* **15**, e0231995, doi:10.1371/journal.pone.0231995 (2020).

44 Jin, J. *et al.* Attention-Block Deep Learning Based Features Fusion in Wearable Social Sensor for Mental Wellbeing Evaluations. *IEEE Access* **8**, 89258-89268, doi:10.1109/ACCESS.2020.2994124 (2020).

45 Khan, N. S., Ghani, M. S. & Anjum, G. ADAM-sense: Anxiety-displaying activities recognition by motion sensors. *Pervasive and Mobile Computing* **78**, 101485, doi:https://doi.org/10.1016/j.pmcj.2021.101485 (2021).

46 Kim, H. *et al.* Depression Prediction by Using Ecological Momentary Assessment, Actiwatch Data, and Machine Learning: Observational Study on Older Adults Living Alone. *JMIR Mhealth Uhealth* **7**, e14149, doi:10.2196/14149 (2019).

47 Kulam, S. *Time-Series Classification with Uni-Dimensional Convolutional Neural Networks: An Experimental Comparison with Long Short-Term Memory Networks*, (2019).

48 Kumar, A., Sangwan, S. R., Arora, A. & Menon, V. G. Depress-DCNF: A deep convolutional neuro-fuzzy model for detection of depression episodes using IoMT. *Applied Soft Computing* **122**, 108863, doi:https://doi.org/10.1016/j.asoc.2022.108863 (2022).

49 Llamocca, P., López, V., Santos, M. & Čukić, M. Personalized Characterization of Emotional States in Patients with Bipolar Disorder. *Mathematics* **9**, doi:10.3390/math9111174 (2021).

50 Lu, J. *et al.* Joint Modeling of Heterogeneous Sensing Data for Depression Assessment via Multi-task Learning. *Proc. ACM Interact. Mob. Wearable Ubiquitous*

Technol. **2**, Article 21, doi:10.1145/3191753 (2018).
51 Mahendran, N. *et al.* Sensor-Assisted Weighted Average Ensemble Model for Detecting Major Depressive Disorder. *Sensors (Basel)* **19**, doi:10.3390/s19224822 (2019).
52 Makhmutova, M. *et al.* Predicting Changes in Depression Severity Using the PSYCHE-D (Prediction of Severity Change-Depression) Model Involving Person-Generated Health Data: Longitudinal Case-Control Observational Study. *JMIR Mhealth Uhealth* **10**, e34148, doi:10.2196/34148 (2022).
53 Mallikarjun, H. M. & Manimegalai, P. Manoglanistara - Emotional Wellness Phases Prediction of Adolescent Female Students by using Brain Waves. *Current Signal Transduction Therapy* **15**, 315-323, doi:10.2174/1574362414666190703151853 (2020).
54 McGinnis, E. W. *et al.* Digital Phenotype for Childhood Internalizing Disorders: Less Positive Play and Promise for a Brief Assessment Battery. *IEEE J Biomed Health Inform* **25**, 3176-3184, doi:10.1109/jbhi.2021.3053846 (2021).
55 McGinnis, R. S. *et al.* Rapid Anxiety and Depression Diagnosis in Young Children Enabled by Wearable Sensors and Machine Learning. *Annu Int Conf IEEE Eng Med Biol Soc* **2018**, 3983-3986, doi:10.1109/embc.2018.8513327 (2018).
56 McGinnis, R. S. *et al.* Rapid detection of internalizing diagnosis in young children enabled by wearable sensors and machine learning. *PLoS One* **14**, e0210267, doi:10.1371/journal.pone.0210267 (2019).
57 Minaeva, O. *et al.* Screening for Depression in Daily Life: Development and External Validation of a Prediction Model Based on Actigraphy and Experience Sampling Method. *J Med Internet Res* **22**, e22634, doi:10.2196/22634 (2020).
58 Miranda, D., Favela, J. & Arnrich, B. Detecting Anxiety States when Caring for People with Dementia. *Methods Inf Med* **56**, 55-62, doi:10.3414/me15-02-0012 (2017).
59 Mullick, T., Radovic, A., Shaaban, S. & Doryab, A. Predicting Depression in Adolescents Using Mobile and Wearable Sensors: Multimodal Machine Learning-Based Exploratory Study. *JMIR Form Res* **6**, e35807, doi:10.2196/35807 (2022).
60 Narziev, N. *et al.* STDD: Short-Term Depression Detection with Passive Sensing. *Sensors (Basel)* **20**, doi:10.3390/s20051396 (2020).
61 Nath, R. K. & Thapliyal, H. Machine Learning-Based Anxiety Detection in Older Adults Using Wristband Sensors and Context Feature. *SN Computer Science* **2**, 359, doi:10.1007/s42979-021-00744-z (2021).
62 Nguyen, D.-K., Chan, C.-L., Li, A.-H. A. & Phan, D.-V. in *2021 5th International Conference on Medical and Health Informatics* 7–12 (Association for Computing Machinery, Kyoto, Japan, 2021).
63 Nishimura, Y. *et al.* in *Sensor-and Video-Based Activity and Behavior Computing* 1-26 (Springer, 2022).
64 Opoku Asare, K. *et al.* Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis. *Pervasive and Mobile Computing* **83**, 101621, doi:https://doi.org/10.1016/j.pmcj.2022.101621 (2022).
65 Pacheco-González, S. L. *et al.* Evaluation of Five Classifiers for Depression Episodes Detection. *Res. Comput. Sci.* **148**, 129-138, doi:10.13053/rcs-148-10-11 (2019).
66 Pedrelli, P. *et al.* Monitoring Changes in Depression Severity Using Wearable and Mobile Sensors. *Front Psychiatry* **11**, 584711, doi:10.3389/fpsyt.2020.584711 (2020).
67 Qian, K. *et al.* in *Proceedings of the Third International Symposium on Image*

Computing and Digital Medicine 71–75 (Association for Computing Machinery, Xi'an, China, 2019).

68 Raihan, M., Bairagi, A. K. & Rahman, S. in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. 1-5.

69 Rodríguez-Ruiz, J. G. *et al.* Classification of Depressive and Schizophrenic Episodes Using Night-Time Motor Activity Signal. *Healthcare (Basel)* **10**, doi:10.3390/healthcare10071256 (2022).

70 Rodríguez-Ruiz, J. G. *et al.* Classification of Depressive Episodes Using Nighttime Data; a Multivariate and Univariate Analysis. *Program. Comput. Softw.* **46**, 689–698, doi:10.1134/s0361768820080198 (2020).

71 Rodríguez-Ruiz, J. G. *et al.* Comparison of Night, Day and 24 h Motor Activity Data for the Classification of Depressive Episodes. *Diagnostics (Basel)* **10**, doi:10.3390/diagnostics10030162 (2020).

72 Rother, R., Sun, Y. & Lo, B. in *Living in the Internet of Things (IoT 2019)*. 1-6.

73 Rykov, Y. *et al.* Digital Biomarkers for Depression Screening With Wearable Devices: Cross-sectional Study With Machine Learning Modeling. *JMIR Mhealth Uhealth* **9**, e24872, doi:10.2196/24872 (2021).

74 Saha, K. *et al.* Person-Centered Predictions of Psychological Constructs with Social Media Contextualized by Multimodal Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **5**, Article 32, doi:10.1145/3448117 (2021).

75 Šalkevicius, J., Damaševičius, R., Maskeliunas, R. & Laukienė, I. Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals. *Electronics* **8**, doi:10.3390/electronics8091039 (2019).

76 Scism, J. E. *Applications of Wearable Sensors in Delivering Biologically Relevant Signals*. (The University of Vermont and State Agricultural College, 2020).

77 Shah, R. V. *et al.* Personalized machine learning of depressed mood using wearables. *Translational Psychiatry* **11**, 338, doi:10.1038/s41398-021-01445-0 (2021).

78 Shaukat-Jali, R., van Zalk, N. & Boyle, D. E. Detecting Subclinical Social Anxiety Using Physiological Data From a Wrist-Worn Wearable: Small-Scale Feasibility Study. *JMIR Form Res* **5**, e32656, doi:10.2196/32656 (2021).

79 Tazawa, Y. *et al.* Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. *Heliyon* **6**, e03274, doi:10.1016/j.heliyon.2020.e03274 (2020).

80 Tiwari, A., Cassani, R., Narayanan, S. & Falk, T. H. in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2213-2216.

81 Tsai, C. H. *et al.* Panic Attack Prediction Using Wearable Devices and Machine Learning: Development and Cohort Study. *JMIR Med Inform* **10**, e33063, doi:10.2196/33063 (2022).

82 Valenza, G. *et al.* Characterization of depressive States in bipolar patients using wearable textile technology and instantaneous heart rate variability assessment. *IEEE J Biomed Health Inform* **19**, 263-274, doi:10.1109/jbhi.2014.2307584 (2015).

83 Wang, R. *et al.* Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2**, Article 43, doi:10.1145/3191775 (2018).

84 Xu, X. *et al.* Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **3**, Article 116, doi:10.1145/3351274 (2019).

85 Yadav, M. *et al.* in *2019 8th International Conference on Affective Computing and*

Intelligent Interaction (ACII). 1-7.

86 Zanella-Calzada, L. A. *et al.* Feature Extraction in Motor Activity Signal: Towards a Depression Episodes Detection in Unipolar and Bipolar Patients. *Diagnostics (Basel)* **9**, doi:10.3390/diagnostics9010008 (2019).

87 Zheng, Y., Wong, T. C. H., Leung, B. H. K. & Poon, C. C. Y. Unobtrusive and Multimodal Wearable Sensing to Quantify Anxiety. *IEEE Sensors Journal* **16**, 3689-3696, doi:10.1109/JSEN.2016.2539383 (2016).

88 Kang, M. & Chai, K. Wearable Sensing Systems for Monitoring Mental Health. *Sensors (Basel)* **22**, doi:10.3390/s22030994 (2022).

89 Byeon, H. Relationship between Physical Activity Level and Depression of Elderly People Living Alone. *International Journal of Environmental Research and Public Health* **16**, doi:10.3390/ijerph16204051 (2019).

90 Puccinelli, P. J. *et al.* Reduced level of physical activity during COVID-19 pandemic is associated with depression and anxiety levels: an internet-based survey. *BMC Public Health* **21**, 425, doi:10.1186/s12889-021-10470-z (2021).

91 Chalmers, J. A., Quintana, D. S., Abbott, M. J. & Kemp, A. H. Anxiety Disorders are Associated with Reduced Heart Rate Variability: A Meta-Analysis. *Front Psychiatry* **5**, 80, doi:10.3389/fpsyt.2014.00080 (2014).

92 Nelson, B. W. & Allen, N. B. Accuracy of Consumer Wearable Heart Rate Measurement During an Ecologically Valid 24-Hour Period: Intraindividual Validation Study. *JMIR Mhealth Uhealth* **7**, e10828, doi:10.2196/10828 (2019).

93 Bent, B., Goldstein, B. A., Kibbe, W. A. & Dunn, J. P. Investigating sources of inaccuracy in wearable optical heart rate sensors. *npj Digital Medicine* **3**, 18, doi:10.1038/s41746-020-0226-6 (2020).

94 Hunkin, H., King, D. L. & Zajac, I. T. Perceived acceptability of wearable devices for the treatment of mental health problems. *Journal of Clinical Psychology* **76**, 987-1003, doi:https://doi.org/10.1002/jclp.22934 (2020).

95 Lovato, N. & Gradisar, M. A meta-analysis and model of the relationship between sleep and depression in adolescents: Recommendations for future research and clinical practice. *Sleep Medicine Reviews* **18**, 521-529, doi:https://doi.org/10.1016/j.smrv.2014.03.006 (2014).

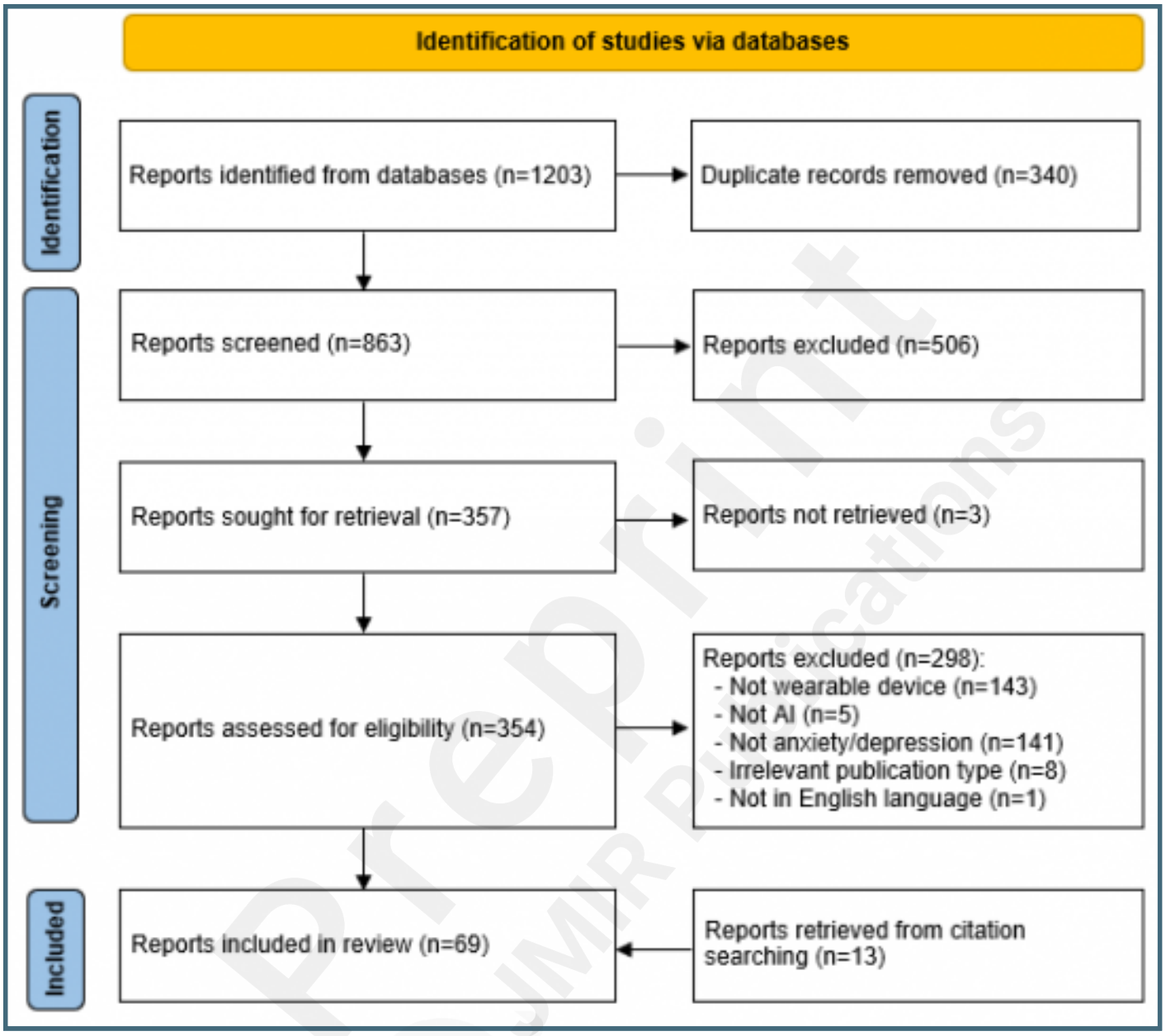
96 Ramsawh, H. J. *et al.* Relationship of anxiety disorders, sleep quality, and functional impairment in a community sample. *Journal of Psychiatric Research* **43**, 926-933, doi:https://doi.org/10.1016/j.jpsychires.2009.01.009 (2009).

97 Murnane, E. L. *et al.* Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. *J Am Med Inform Assoc* **23**, 477-484, doi:10.1093/jamia/ocv165 (2016).

Supplementary Files

Figures

Flow diagram of the study selection process.



Multimedia Appendixes

PRISMA-ScR-Checklist.
URL: http://asset.jmir.pub/assets/77b61dce3c579c40040cd653f39aff39.docx

Search strategy.
URL: http://asset.jmir.pub/assets/bc374c2fa62b871a34d98d17f063ef20.docx

Data extraction form.
URL: http://asset.jmir.pub/assets/8418ad8524037feaa521e10ab3a9c64d.docx

Characteristics of each included study.
URL: http://asset.jmir.pub/assets/3f01ff107f4206b3776f85a23df4973b.docx

Features of wearable devices.
URL: http://asset.jmir.pub/assets/2af9c55371f7dc4c3824040079c1d678.docx

Features of sensors of wearable devices.
URL: http://asset.jmir.pub/assets/b6019d0fdc77e4487925a8c9a706169b.docx

Features of AI algorithms.
URL: http://asset.jmir.pub/assets/736dc9d71b783e86327f4095679c0e4f.docx

Features of data used in AI algorithms.
URL: http://asset.jmir.pub/assets/ba5bcb2172e9504de8f3fa68527a02d0.docx

TOC/Feature image for homepages

Untitled.