

## Review

# The Performance of Wearable AI in Detecting Stress Among Students: Systematic Review and Meta-Analysis

Alaa Abd-alrazaq<sup>1</sup>, PhD; Mohammad Alajlani<sup>2</sup>, PhD; Reham Ahmad<sup>2</sup>, MSc; Rawan AlSaad<sup>1</sup>, PhD; Sarah Aziz<sup>1</sup>, MSc; Arfan Ahmed<sup>1</sup>, PhD; Mohammed Alsahli<sup>3</sup>, PhD; Rafat Damseh<sup>4</sup>, PhD; Javaid Sheikh<sup>1</sup>, MD

<sup>1</sup>AI Center for Precision Health, Weill Cornell Medicine-Qatar, Qatar Foundation, Doha, Qatar

<sup>2</sup>Institute of Digital Healthcare, WMG, University of Warwick, Warwick, United Kingdom

<sup>3</sup>Health Informatics Department, College of Health Science, Saudi Electronic University, Riyadh, Saudi Arabia

<sup>4</sup>Department of Computer Science and Software Engineering, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates

**Corresponding Author:**

Alaa Abd-alrazaq, PhD

AI Center for Precision Health

Weill Cornell Medicine-Qatar

Qatar Foundation

PO Box 5825, Doha Al Luqta St, Ar-Rayyan

Doha, 2442

Qatar

Phone: 974 5570845212

Email: [Aaa4027@qatar-med.cornell.edu](mailto:Aaa4027@qatar-med.cornell.edu)

## Abstract

**Background:** Students usually encounter stress throughout their academic path. Ongoing stressors may lead to chronic stress, adversely affecting their physical and mental well-being. Thus, early detection and monitoring of stress among students are crucial. Wearable artificial intelligence (AI) has emerged as a valuable tool for this purpose. It offers an objective, noninvasive, nonobtrusive, automated approach to continuously monitor biomarkers in real time, thereby addressing the limitations of traditional approaches such as self-reported questionnaires.

**Objective:** This systematic review and meta-analysis aim to assess the performance of wearable AI in detecting and predicting stress among students.

**Methods:** Search sources in this review included 7 electronic databases (MEDLINE, Embase, PsycINFO, ACM Digital Library, Scopus, IEEE Xplore, and Google Scholar). We also checked the reference lists of the included studies and checked studies that cited the included studies. The search was conducted on June 12, 2023. This review included research articles centered on the creation or application of AI algorithms for the detection or prediction of stress among students using data from wearable devices. In total, 2 independent reviewers performed study selection, data extraction, and risk-of-bias assessment. The Quality Assessment of Diagnostic Accuracy Studies-Revised tool was adapted and used to examine the risk of bias in the included studies. Evidence synthesis was conducted using narrative and statistical techniques.

**Results:** This review included 5.8% (19/327) of the studies retrieved from the search sources. A meta-analysis of 37 accuracy estimates derived from 32% (6/19) of the studies revealed a pooled mean accuracy of 0.856 (95% CI 0.70-0.93). Subgroup analyses demonstrated that the accuracy of wearable AI was moderated by the number of stress classes ( $P=.02$ ), type of wearable device ( $P=.049$ ), location of the wearable device ( $P=.02$ ), data set size ( $P=.009$ ), and ground truth ( $P=.001$ ). The average estimates of sensitivity, specificity, and  $F_1$ -score were 0.755 (SD 0.181), 0.744 (SD 0.147), and 0.759 (SD 0.139), respectively.

**Conclusions:** Wearable AI shows promise in detecting student stress but currently has suboptimal performance. The results of the subgroup analyses should be carefully interpreted given that many of these findings may be due to other confounding factors rather than the underlying grouping characteristics. Thus, wearable AI should be used alongside other assessments (eg, clinical questionnaires) until further evidence is available. Future research should explore the ability of wearable AI to differentiate types of stress, distinguish stress from other mental health issues, predict future occurrences of stress, consider factors such as the placement of the wearable device and the methods used to assess the ground truth, and report detailed results to facilitate the conduct of meta-analyses.

**Trial Registration:** PROSPERO CRD42023435051; <http://tinyurl.com/3fzb5rnp>

**KEYWORDS**

stress; artificial intelligence; wearable devices; machine learning; systematic review; students; mobile phone

## Introduction

### Background

Students' mental health is now of significant concern in the domain of public health. The academic environment, composed of rigorous curriculum structures, exams, assignments, evaluative metrics, deadlines, and an undercurrent of peer comparisons, not only requires a high cognitive effort but can also be a basis for psychosocial risks [1-3]. Being immersed in such scholarly endeavors at different levels, students become inadvertently vulnerable to mental health issues (eg, anxiety, depression, and stress [3,4]). This can negatively affect a student's life in general and, specifically, their academic performance. Research has revealed that students who are constantly under high levels of stress are more inclined to underperform academically when compared with their peers who are less stressed [5]. Recent studies suggest that failure to manage academic-related stress is detrimental to the well-being of students [6-8].

The psychobiological literature defines stress, including academic stress, as a syndrome of responses combining interactions between the neurocognitive, endocrine, and affective systems [9-11]. Stress manifests when exposed to conditions that present a physical and psychological challenge or threat [11]. Stress is also defined as the result of incompatibility in relation to an individual's ability to handle the burden placed on them by their environment [12]. Stress can be classified based on its duration into 2 forms: chronic and acute stress. Acute stress is a short-term stress response to an immediate situation and usually subsides soon after symptoms first appear, whereas chronic stress is a long-term stress response to ongoing stressors [13]. Typically, acute stress is the most common and does not impose a health burden on young and healthy individuals. In contrast, chronic stress can have a negative impact on both physical and mental health [14,15]. Potential biological risks include inflammation, diabetes, and heart disease [16,17]. Furthermore, studies have shown an association with cognitive dysfunctions because of excessive and chronic stressors [18,19]. Chronic stress can also have negative effects on social behavior, which leads to antisociality [20]; depression; or, in extreme cases, suicide [21,22]. Worldwide, it has been reported that, for every 5 visits by college students to their physician, 3 are for stress-related problems [23]. In addition, most college students who have problems with stress also have sleep-related disorders (76%), are prone to headaches (58%), experience difficult relationships with family and friends (85%), and tend to be short-tempered (70%) [23]. The American Institute of Stress reveals that, in the United States, a combined US \$300 billion is spent on treating stress-related ailments and diseases [13].

Given the aforementioned complications of stress, the detection and monitoring of stress in early stages are crucial. Subjective and objective assessments can be used to evaluate stress levels.

Subjective assessments of stress can involve 2 different means: standard self-reported questionnaires designed by field experts (eg, the Perceived Stress Scale and Depression, Anxiety, and Stress Scale–21 Items) or clinical interviews with psychologists [15,24]. Beyond this, objective assessments of stress include physical observations and physiological measures. For physical observations, visible changes in the appearance of the human body are noted, such as any facial expressions, the rate of blinking, and any noticeable dilation of the pupils. For physiological measures, blood, urine, or saliva samples can be collected to measure the level of specific hormones (adrenaline and cortisol) that are released as a response to a stressor [10,14,16]. Relying only on subjective assessments is insufficient to accurately detect and monitor stress as they are highly subjective, usually irreproducible, time-consuming, and entangled with other body responses not related to stress [25,26]. Although the aforementioned objective assessments can be deterministic in capturing stress status, they require invasive or obtrusive tests that are commonly deployed in a highly controlled laboratory environment and do not reflect stress in real-life scenarios [13,27]. Therefore, there is a critical need for alternative assessments of stress that are objective, noninvasive, nonobtrusive, and automatic and can continuously monitor biomarkers in real time.

Changes in biomarkers (eg, heart rate [HR], HR variability [HRV], and electrodermal activity [EDA]) can be used to objectively detect stress [15,28]. Real-time and continuous monitoring of these biomarkers must be supported by computational frameworks to automate the process of stress detection, thereby overcoming the aforementioned limitations of the current tools. Wearable artificial intelligence (AI) is a promising solution that has been used to address the issue of assessing stress [29-31]. Wearable AI is an advanced technology that depends on AI techniques to analyze a large amount of data (eg, HR, HRV, EDA, activity level, and skin temperature) collected by sensors in wearable devices to provide personalized feedback. Several types of wearable devices can be used to collect biomarkers: on-body devices, which are fixed directly on the body or skin; near-body devices, which are fixed close to the body but do not have direct contact with the skin; in-body devices, which are implanted in the body; and electronic textiles, which are fabrics that have integrated electronics [30].

### Research Problem and Aim

In the past few years, numerous studies have shed light on the performance of wearable AI for the detection of stress. Several reviews of such studies have been conducted, but they have the following limitations. First, none of the previous reviews have focused on students [13,15,24,28,32-35]. Second, none of the previous reviews have used statistical techniques (eg, meta-analysis) to analyze the results of previous studies [13,15,24,28,32-35]. Third, most of the previous reviews have been literature reviews rather than systematic reviews [13,15,24,28,33,34]. Fourth, the search sources in the previous

reviews have not included the main databases in the field, such as MEDLINE [28], ACM Digital Library [28,32,33,35], Scopus [28,32,33,35], Embase [32,35], and IEEE Xplore [32,33,35]. Finally, several previous reviews have focused on wearable and nonwearable devices instead of on wearable devices only [13,15,24,32]. Therefore, this review sought to bridge this gap by assessing the capability of wearable AI to detect and forecast stress among students. Our research question is as follows: what is the effectiveness of wearable AI in diagnosing and predicting stress among students?

## Methods

### Overview

The authors undertook and reported this review in line with the guidelines set forth by the PRISMA-DTA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Diagnostic Test Accuracy) [36]. The PRISMA-DTA checklist for this review is presented in [Multimedia Appendix 1](#) [36]. The protocol for this review was registered with PROSPERO (ID: CRD42023435051).

### Search Strategy

To identify suitable studies, the primary author conducted a comprehensive search across 7 electronic repositories on June 12, 2023: MEDLINE (via Ovid), Embase (via Ovid), PsycINFO (via Ovid), ACM Digital Library, Scopus, IEEE Xplore, and Google Scholar. Using an automated search methodology, biweekly notifications were configured over a span of 3 months, concluding on September 12, 2023. Given the large number of results returned by Google Scholar, the first 100 results (equivalent to 10 pages) were subjected to scrutiny for this review. To find more studies, we checked the lists of sources referenced in the studies that we had already included (backward reference list checking), and we also looked at studies that had cited the ones we had already included (forward reference list checking).

The compilation of search terms for the review was informed by consultations with 2 digital mental health experts and scrutiny of relevant literature reviews. The resultant search query was composed of four distinct categories of terms: (1) terms associated with AI, such as “artificial intelligence,” “machine learning,” and “deep learning”; (2) terms linked to wearable devices, comprising “wearable\*,” “smartwatch\*,” and “smartband\*”; (3) terms related to stress; and (4) terms relevant to students, including “student\*,” “postgraduate\*,” and “undergraduate\*.” The Boolean OR operator was used to combine terms in the same category, whereas the Boolean AND operator was used to combine terms between categories. The precise search formulations are presented in [Multimedia Appendix 2](#) for reference.

### Study Eligibility Criteria

This review analyzed research articles centered on the creation or application of AI algorithms for the detection or prediction of stress among students using data from wearable devices. The standards for determining which articles to include and exclude were jointly established through the collaborative expertise of the authors. To be eligible for consideration in this review,

studies needed to collect data from students regardless of their educational level, age, gender, and ethnicity. Furthermore, to be included in this review, studies had to assess the effectiveness of AI algorithms in identifying or foreseeing stress. They were also required to present the confusion matrix or performance metrics (such as accuracy, sensitivity, and specificity). Studies that used AI to predict the outcomes of stress interventions or treatments were excluded.

Eligible studies were required to use noninvasive wearable devices (eg, smartwatches, smart glasses, and smart rings) for data collection. This review encompassed studies that used various methods for data collection (such as nonwearable devices, interviews, and questionnaires) in addition to wearable devices. In contrast, studies that solely used the following devices for data collection were excluded: nonwearable devices, handheld devices (eg, mobile phones), near-body wearable devices, in-body wearable devices, wearable devices wired to nonwearable devices, and wearables requiring expert oversight (such as those requiring precise electrode placement).

The inclusion criteria covered peer-reviewed journal articles, conference papers, preprints, and dissertations irrespective of the study setting, reference standard (ie, ground truth), or the country of study. Given the emphasis on contemporary technology and the continuous growth in the field of wearable AI, only articles from 2015 onward were included. Articles not in English or structured as review pieces, editorials, conference abstracts, posters, protocols, and research highlights were all excluded. We included only prospective and retrospective experimental studies; however, studies that used publicly available data sets (eg, Wearable Stress and Affect Detection [WESAD]) or data sets from previous studies were excluded.

### Study Selection

The study selection process involved 3 main steps. First, duplicate studies were removed using EndNote X9 (Clarivate Analytics). Then, 2 reviewers independently evaluated the titles and abstracts of the remaining articles. Finally, the reviewers independently assessed the full texts of the remaining articles. Any disagreements were discussed and resolved. The agreement between the reviewers was high, with a  $\kappa$  score of 0.87 for evaluating titles and abstracts and 0.92 for reading full texts.

### Data Extraction

The data extraction form was developed and tested with 3 studies first ([Multimedia Appendix 3](#)). Using Microsoft Excel (Microsoft Corp), 2 reviewers independently extracted the studies' metadata, features of the wearable devices, features of the AI algorithms, and the study results. Any disagreements between the reviewers were resolved through discussion. For studies that provided raw data or confusion matrices, we calculated performance measures such as accuracy, specificity, and sensitivity. If this information was not available, we reached out to the first and corresponding authors of those studies to try to obtain it. We did not include results based only on nonwearable device data (eg, data from smartphones or surveys). As many studies conducted multiple experiments with different features, data types, validation methods, and AI algorithms, they reported several results for the same performance measure.

In these cases, we obtained the best result in each performance measure for each algorithm.

### Risk-of-Bias and Applicability Appraisal

To assess the quality of the included studies, we tailored the established Quality Assessment of Diagnostic Accuracy Studies–Revised (QUADAS-2) [37] tool to suit the objectives of our review. We made adjustments by replacing certain nonapplicable criteria with more relevant ones from the Prediction Model Risk of Bias Assessment Tool [38]. Our adjusted QUADAS-2 tool includes 4 primary categories: “Participants,” “Index Test” (pertaining to AI algorithms), “Reference Standard” (denoting the ground truth), and “Analysis.” We formulated 4 specific questions within each category to align with the aims of our review. In addition to examining the potential biases within these 4 categories, we assessed the real-world applicability of the findings for the first 3 categories. To optimize our adjusted tool, we first trialed it on 3 studies to fine-tune it. In total, 2 reviewers independently assessed the included studies using the adapted QUADAS-2 tool (Multimedia Appendix 4). Discrepancies between their evaluations were discussed and resolved through consensus.

### Data Synthesis

The data from the studies were combined using narrative and statistical methods. In the narrative synthesis, we used text and tables to provide an overview and describe the key features of the included studies (study metadata, wearable devices, and AI techniques). A statistical approach was used when at least 2 different studies presented sufficient data to conduct meta-analyses. We did not consider the study design when selecting studies for meta-analysis. When estimates (eg, accuracy, sensitivity, and specificity) in the analysis were extracted from different unique studies (ie, independent effect sizes), DerSimonian-Laird random-effects models [39] using the Freeman-Tukey double arcsine transformation [40,41] were conducted to pool the extracted estimates. This approach considers the fluctuations arising from sampling and the heterogeneity in estimates. The analysis was executed using the *meta* toolkit in R (version 4.2.2; R Foundation for Statistical Computing) [42]. However, when there were estimates in the analysis extracted from the same study (ie, dependent effect

sizes), we used a multilevel meta-analysis technique [39,43] to account for this dependency in effect sizes, thereby reducing the likelihood of type-I errors. Multilevel meta-analyses were conducted using the *metafor* toolkit in R (version 4.2.2) [40].

Where appropriate, we performed subgroup multilevel meta-analyses to investigate potential relationships between performance estimates and various factors [39,43]. These factors included AI algorithms, number of stress classes, type of wearable device, location of the wearable device, data set size, data sources, data types, stress inducers, ground truth, and validation methods. The difference in estimates between subgroups was considered statistically significant when the *P* value was  $<.05$ .

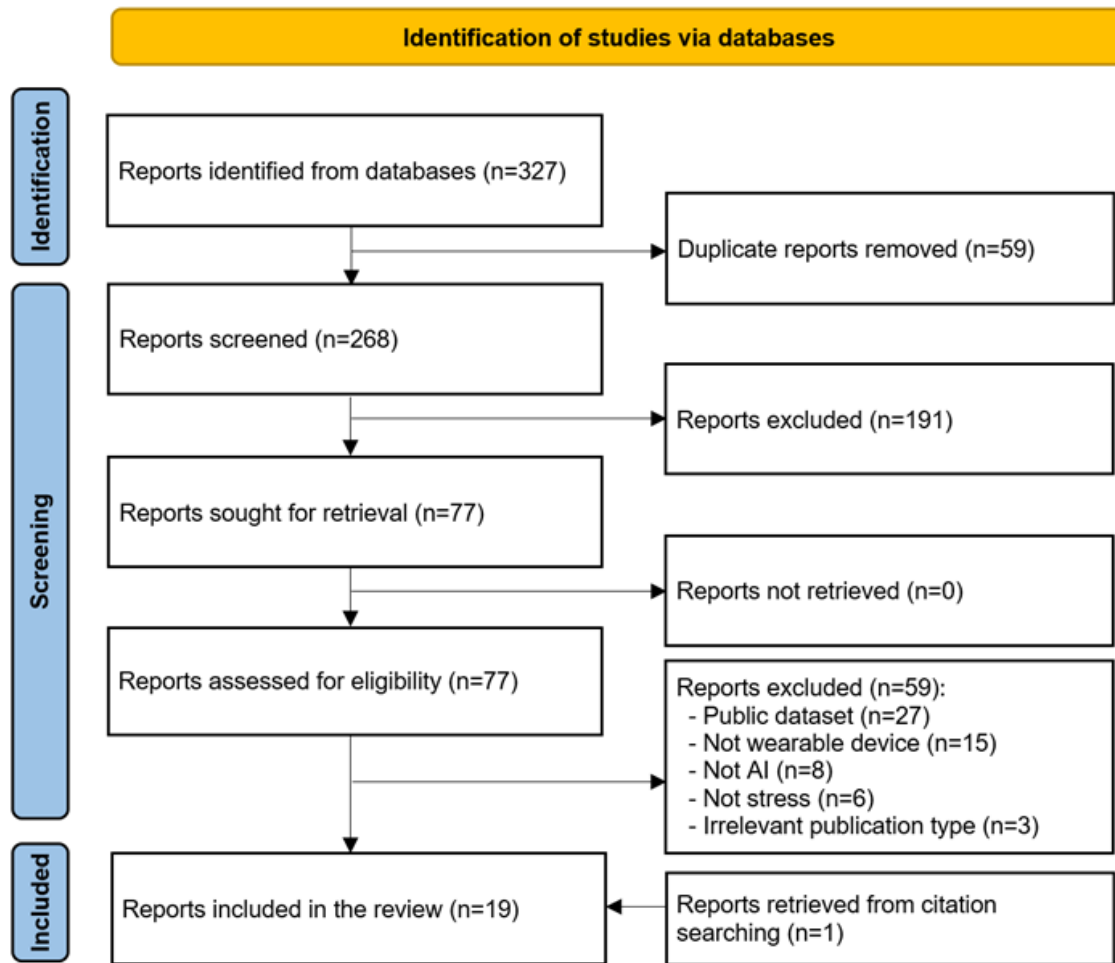
The Cochrane *Q* statistic was used to examine between-study heterogeneity, with a *P* value of  $<.05$  indicating the presence of heterogeneity. The degree of between-study heterogeneity was evaluated using  $I^2$  [40,44], where it was deemed insignificant when  $I^2$  ranged from 0% to 40%, moderate when it ranged from 30% to 60%, substantial when it ranged from 50% to 90%, or considerable when it ranged from 75% to 100% [45].

## Results

### Search Results

As shown in Figure 1, a total of 327 studies were retrieved when the databases identified previously were searched. Of the 327 retrieved studies, 59 (18%) duplicates were removed using EndNote X9, leaving 268 (82%) studies. Furthermore, 71.3% (191/268) of the studies were removed after we screened the titles and abstracts. After retrieving and reading the full texts of the remaining 77 studies, it was determined that 59 (77%) were ineligible for inclusion. The main reasons for exclusion were that they used a public data set, did not use wearable devices, did not use AI algorithms, did not focus on stress, or were irrelevant publication types. We identified an additional study relevant to this review through backward reference list checking. In total, 19 studies were included in this review [46-64], and 6 (32%) were eligible for the meta-analyses [46,52,53,55,62,63].

Figure 1. Flowchart of the study selection process. AI: artificial intelligence.



### Characteristics of the Included Studies

The key characteristics of the studies included in this review are presented in Table 1. The included studies were published between 2015 and 2023. The year in which the largest number of included studies was published was 2019 (5/19, 26%). The included studies were conducted in 10 distinct countries (Table 1), wherein the United States contributed to over a third of the studies (7/19, 37%). More than half (10/19, 53%) of the studies were conference papers, whereas the remaining studies were journal articles (8/19, 42%) and a preprint (1/19, 5%). The

number of participants in the included studies ranged from 5 to 652, with an average of 77 (SD 149.5). The mean age of the participants was reported in 42% (8/19) of the included studies and ranged from 19.8 to 27.5 years, with an average of 22.8 (SD 2.5) years. A total of 58% (11/19) of the studies reported the proportion of female participants, which ranged from 14.3% to 76.8%, with an average of 47.3% (SD 21.5%). Participants in most of the studies (17/19, 89%) were undergraduate students. The characteristics of each included study are presented in Multimedia Appendix 5 [46-64].

**Table 1.** Characteristics of the included studies (N=19).

Feature	Values	References
<b>Year of publication, n (%)</b>		
2023	2 (11)	[47,54]
2021	3 (16)	[46,53,63]
2020	4 (21)	[49,56,62,64]
2019	5 (26)	[48,52,55,57,58]
2018	3 (16)	[50,60,61]
2017	1 (5)	[51]
2015	1 (5)	[59]
<b>Publication type, n (%)</b>		
Conference paper	10 (53)	[46,51-53,55,57-59,61,64]
Journal article	8 (42)	[48-50,54,56,60,62,63]
Preprint	1 (5)	[47]
<b>Country of publication, n (%)</b>		
United States	7 (37)	[47,51,53-56,60]
India	3 (16)	[57,58,63]
United Kingdom	2 (11)	[49,59]
Others	1 (5) each	[46,48,50,52,61,62,64]
Number of participants, mean (SD; range)	77 (149.5; 5-652)	[46-64]
<b>Age (y)</b>		
Values, mean (SD; range)	22.76 (2.53; 19.8-27.5)	[47-49,54,56,61,62,64]
NR <sup>a</sup> , n (%)	11 (58)	[46,50-53,55,57-60,63]
<b>Female participants</b>		
Values (%), mean (SD; range)	47.3 (21.5; 14.3-76.8)	[47-49,52-54,56,60-62,64]
NR, n (%)	8 (42)	[46,50,51,55,57-59,63]
<b>Educational level, n (%)</b>		
High school	1 (5)	[48]
Undergraduate	17 (89)	[47-60,62-64]
Postgraduate	3 (16)	[49,61,63]
NR	1 (5)	[46]

<sup>a</sup>NR: not reported.

### Features of the Wearable Devices

As depicted in [Table 2](#), although 89% (17/19) of the studies used commercial wearable devices, 5% (1/19) used noncommercial wearable devices. The included studies used 16 different wearable devices. The wearable devices that were used by the largest number of studies were Empatica E4 (3/19, 16%) and Galaxy (Watch Active2 and Gear S, S2, and S3; 3/19,

16%). The most common type of wearable device used in the included studies was smart bands (12/19, 63%). The wearable devices were placed on 7 different parts of the body; however, the wrist was the most common (16/19, 84%). The periods for which the devices were worn in the included studies ranged from 10 minutes to 365 days. The features of the wearable devices in each included study are presented in [Multimedia Appendix 6](#) [46-64].

**Table 2.** Features of the wearable devices (N=19).

Feature	Values	References
<b>Status of WD<sup>a</sup>, n (%)</b>		
Commercial	17 (89)	[46-54,57-60,62-64]
Noncommercial	1 (5)	[51]
Not reported	2 (11)	[55,56]
<b>Name of WD, n (%)</b>		
Empatica E4	3 (16)	[48,61,64]
Galaxy (Watch Active2 and Gear S, S2, and S3)	3 (16)	[47,48,54]
Fitbit (Charge 2)	2 (11)	[53,58]
Microsoft Band 2	2 (11)	[50,62]
Others	1 (5) each	[46,49,51,52,54,59,60,63]
Not reported	2 (11)	[55,56]
<b>Type of WD, n (%)</b>		
Smart bands	12 (63)	[48-51,55,56,58-62,64]
Smartwatches	6 (32)	[46-48,51-54,63]
Electrodes	2 (11)	[52,63]
Others	1 (5) each	[54,57]
<b>Placement of WD, n (%)</b>		
Wrist	16 (84)	[46-51,53-56,58-62,64]
Chest	3 (16)	[51,52,63]
Finger	3 (16)	[54,59,63]
Palm	2 (11)	[52,63]
Others	1 (5) each	[57,63]
<b>Duration of wearing WD</b>		
Range	10 min to 365 d	N/A <sup>b</sup>
Not reported, n (%)	6 (32)	[46,50-52,57,62]

<sup>a</sup>WD: wearable device.

<sup>b</sup>N/A: not applicable.

## Features of the AI

The AI algorithms in most of the included studies (18/19, 95%) were used to solve classification problems (Table 3). Most studies (15/19, 79%) used AI algorithms to categorize the sample into 2 classes (stressed vs not stressed). Among the included studies, 22 different algorithms were used, but the most commonly used algorithms were support vector machine (14/19, 74%) and k-nearest neighbor (13/19, 68%). All the included studies used AI to detect the current stress status, whereas none of them used AI to predict the occurrence of stress in the future. The data set size was reported in 47% (9/19) of the studies and ranged from 30 to 1178, with an average of 517.9 (SD 409). Data were collected using wearable devices (19/19, 100%), self-reported questionnaires (11/19, 58%), and nonwearable devices (eg, smartphones; 6/19, 32%). The included studies used 17 types of data to develop the models. However, the most common data types used to develop the

models were HR data (eg, HR, HRV, and interbeat interval; 16/19, 84%), EDA data (10/19, 53%), and physical activity data (eg, step counts, calories, and metabolic rate) (8/19, 42%). The number of features extracted from the data varied between 4 and 498, with an average of 52.2 (SD 115.1). Stress inducers were used in 63% (12/19) of the studies, and the most common method used to induce stress among participants was arithmetic tasks (7/19, 37%). The ground truth was identified using self-reported questionnaires (eg, Perceived Stress Scale; 11/19, 58%), the context of the experiment (eg, stress exposure period and non-stress exposure period; 9/19, 47%), or unsupervised algorithms (1/19, 5%). The most common method used to validate the performance of the models was k-fold cross-validation (11/19, 58%). The most common measure used in the included studies to evaluate the performance of AI algorithms was accuracy (16/19, 84%). The features of the AI in each included study are described in Multimedia Appendix 7 [46-64].

**Table 3.** Features of the artificial intelligence (AI; N=19).

Feature	Values	References
<b>Problem-solving approaches, n (%)</b>		
Classification	18 (95)	[46-63]
Regression	1 (5)	[50]
Clustering	1 (5)	[64]
<b>Number of stress classes, n (%)</b>		
2	15 (79)	[46,47,49,51,53-62,64]
3	5 (26)	[48,50,52,57,61]
>3	2 (11)	[55,63]
<b>AI algorithms, n (%)</b>		
Support vector machine	14 (74)	[46,48,50-57,59,60,62,63]
K-nearest neighbor	13 (68)	[46,47,50,52-57,61-63]
Random forest	8 (42)	[47,50,51,54,55,61-63]
Logistic regression	6 (32)	[46,48,51,53,62,63]
Decision tree	5 (26)	[50,52,55,58,61]
Naïve Bayes	5 (26)	[46,50,51,58,62]
Artificial neural network	4 (21)	[53,55,62,63]
Multilayer perceptron	4 (21)	[46,48,54,56]
Others	<4	[46-50,52,55,56,60-64]
<b>Aim of AI algorithms, n (%)</b>		
Detection	19 (100)	[46-64]
Data set size, mean (SD; range)	518 (409; 30-1178)	[46,47,52,53,55,61-64]
<b>Data sources, n (%)</b>		
Wearable devices	19 (100)	[46-64]
Self-reported questionnaire	11 (58)	[47,49,51,53-55,58,60,62,63]
Nonwearable devices	6 (32)	[47,53,54,58,60,61]
<b>Data types, n (%)</b>		
Heart rate data	16 (84)	[46-54,56,58,59,61,63,64]
Electrodermal activity data	10 (53)	[48,50-52,54,59-61,63,64]
Activity data	8 (42)	[47,48,53-55,58,60,61]
Sleep data	5 (26)	[53-55,58,60]
Skin temperature	3 (16)	[48,50,60]
Smartphone use data	3 (16)	[47,54,60]
Location	3 (16)	[47,54,60]
Respiratory rate data	3 (16)	[47,54,61]
Others	<3	[47,53,54,57,58,60,61]
<b>Number of features</b>		
Values, mean (SD; range)	52.2 (115.1; 4-498)	[46-64]
Not reported, n (%)	1 (5)	[55]
<b>Stress-induction methods, n (%)</b>		
Arithmetic tasks	7 (37)	[46,50,51,56,59,61,63]
Exams	4 (21)	[49,50,52,62]
Stroop test	4 (21)	[50,51,56,63]



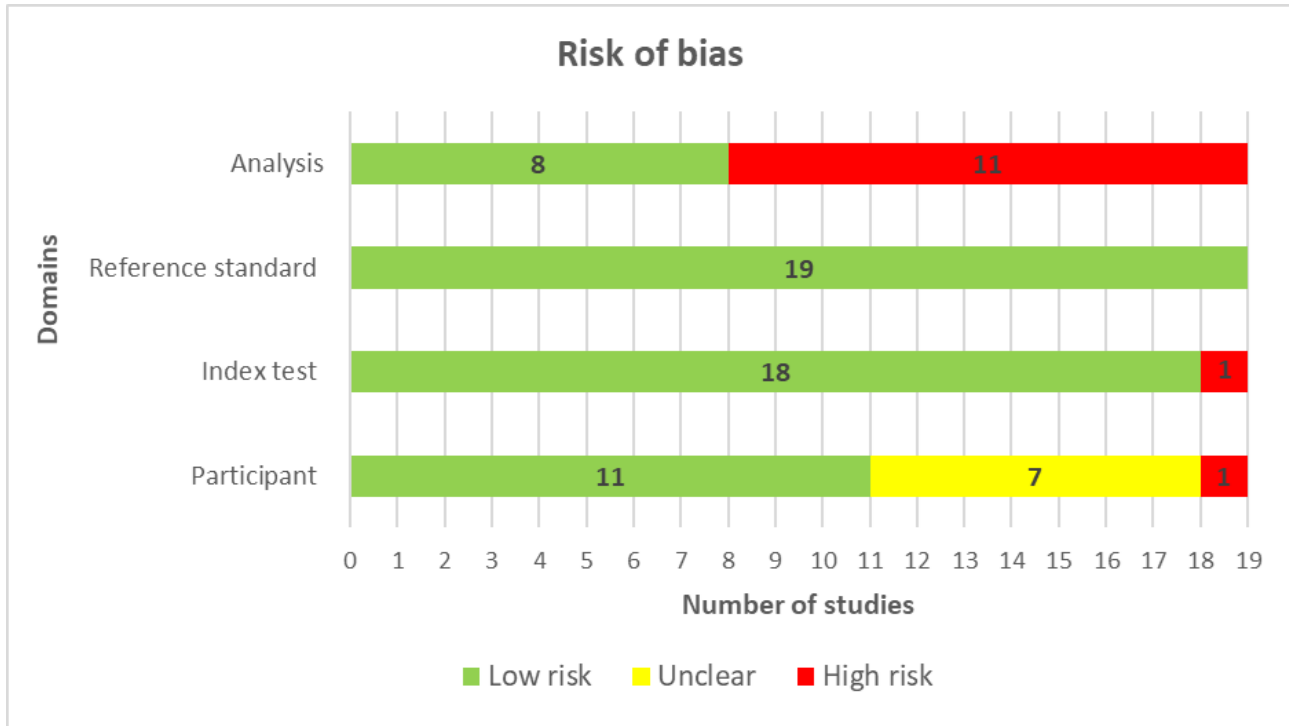
Feature	Values	References
Physical stress	3 (16)	[50,51,56]
Others	<3	[48,50-52,56,59,61,64]
No stress induction	7 (37)	[47,53-55,57,58,60]
<b>Ground truth, n (%)</b>		
Self-reported questionnaire	11 (58)	[47-49,51,53-55,57,58,60,62]
Context	9 (47)	[48,50-52,56,59,61,63,64]
Unsupervised algorithm	1 (5)	[46]
<b>Types of validation, n (%)</b>		
K-fold	11 (58)	[46-48,50-54,56,57,63]
Training-test split	6 (32)	[46,49,55,57,59,62]
Leave-one-out cross-validation	2 (11)	[51,61]
Nested	1 (5)	[60]
Not reported	1 (5)	[58]
Not applicable	1 (5)	[64]
<b>Performance measures, n (%)</b>		
Accuracy	16 (84)	[46,48-50,52-61,63,64]
$F_1$ -score	9 (47)	[46-48,51,55,56,60,61,64]
Precision	7 (37)	[48,51,55,56,58,59,64]
Sensitivity	6 (32)	[48,55,56,58,62,64]
$\kappa$	3 (16)	[46,50,51]
Others	<4	[46,50,51,58,62,64]

### Results of the Risk-of-Bias Appraisal

All studies (19/19, 100%) reported comprehensive details to determine whether an appropriate consecutive or random sample of eligible patients was used, and none of the studies had inappropriate exclusions. Nearly half (8/19, 42%) of the studies ensured a balanced number of patients across subgroups. Only a third (6/19, 32%) reported a sufficient sample size, leaving

ambiguity regarding the adequacy of the sample size in most of the studies (13/19, 68%). As a result, only a little over half (11/19, 58%) of the studies were assessed as having a low risk of bias from the “selection of participants” domain (Figure 2). In terms of matching participants to the predefined requirements in the review question, a low level of concern was identified in most of the studies, accounting for 68% (13/19).

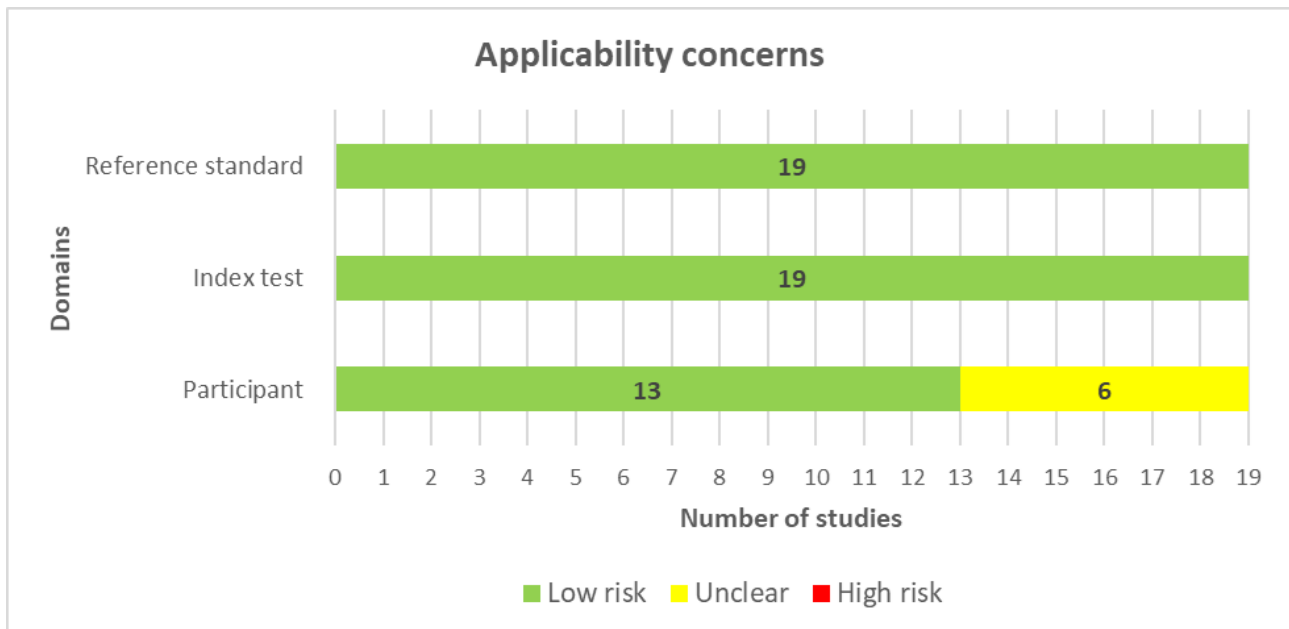
**Figure 2.** Results of the assessment of risk of bias in the included studies.



All studies in our review (19/19, 100%) comprehensively detailed the AI models, clearly reported the features (predictors) used, and ensured that these features were sourced without previous knowledge of the outcome data. For nearly every study (18/19, 95%), the features were consistently assessed across participants. Therefore, for the vast majority of the studies

(18/19, 95%), the potential for bias in the “index test” was assessed as low (Figure 2). Consistently, all studies (19/19, 100%) were found to have minimal concerns regarding the match between the model’s predictors and the review question’s criteria (Figure 3).

**Figure 3.** Results of the assessment of applicability concerns in the included studies.



In all the reviewed studies (19/19, 100%), the outcome of interest, specifically the stress level, was consistently assessed using appropriate methodologies. Similarly, outcomes were defined uniformly for participants across studies and determined without access to predictor data, and an appropriate time gap was ensured between the index test and the reference standard. As a result, the potential for bias regarding the “reference

standard” domain was deemed low for all studies (19/19, 100%; Figure 2). In addition, all studies (19/19, 100%) showed minimal concerns over discrepancies between the outcome’s definition, timing, or determination and the review question’s criteria (Figure 3).

Finally, two-thirds of the studies (12/19, 63%) ensured that all the enrolled participants were factored into the data analysis. Over half (10/19, 53%) of these studies executed proper data preprocessing, whereas in 68% (13/19), there was an appropriate split between training, validation, and test sets. The same proportion of studies (13/19, 68%) adopted suitable measures to evaluate the performance of their models. However, considering all these factors, only a minority (8/19, 42%) were deemed to have a minimal “analysis” domain risk of bias (Figure 2). A detailed breakdown of the “risk of bias” and “applicability concerns” for every domain in each study is available in Multimedia Appendix 8 [46-64].

## Results of the Studies

### Overview

Accuracy, sensitivity, specificity, and  $F_1$ -score were reported in 89% (17/19), 32% (6/19), 11% (2/19), and 53% (10/19) of the studies, respectively. Accuracy values were synthesized using both narrative and statistical methods, whereas the values of the remaining measures were synthesized solely through narrative methods. This distinction arises from the fact that 35% (6/17) of the studies that assessed accuracy reported enough data to conduct meta-analyses, whereas none of the studies

evaluating the other measures had sufficient data for such analyses. The results related to accuracy and other measures are presented in the following subsections.

### Accuracy

Wearable AI accuracy was assessed in 89% (17/19) of the studies. From these studies, we extracted the highest accuracy for each algorithm in each study; therefore, we identified 71 accuracy estimates. The estimates of accuracy ranged from 0.48 to 1.00, with an average of 0.822 (SD 0.135). We conducted meta-analyses of 37 estimates derived from 11,425 participants across 35% (6/17) of the studies (Table 4). The pooled mean accuracy of these estimates was 0.856 (95% CI 0.698-0.934). The meta-analyzed evidence exhibited considerable statistical heterogeneity ( $P<.001$ ;  $I^2=98.8\%$ ). Furthermore, subgroup meta-analyses were conducted to assess the performance of wearable AI based on different factors. As shown in Table 4, there was a statistically significant difference in pooled accuracy between subgroups in the number of stress classes group ( $P=.02$ ), type of wearable device group ( $P=.049$ ), location of wearable device group ( $P=.02$ ), data set size group ( $P=.009$ ), and ground truth group ( $P=.001$ ), whereas no statistically significant difference was found in the pooled accuracy between subgroups in the remaining groups.

**Table 4.** Pooled mean estimates of accuracy by several factors (N=19).

Group	Number of studies <sup>a</sup>	Sample size, n	Accuracy (%), range	Pooled mean accuracy (%), mean (95% CI)	Heterogeneity measures			Test for subgroup differences (P value)
					$\tau^2$	Q (P value)	$I^2$ (%)	
<b>Algorithm</b>								.98
Support vector machine	6	2039	0.65-0.94	0.796 (0.67-0.90)	0.0276	143.0 (<.001)	96.5	
K-nearest neighbor	6	2039	0.58-0.98	0.834 (0.71-0.93)	0.0326	96.2 (<.001)	94.8	
Logistic regression	4	1545	0.66-0.87	0.768 (0.67-0.86)	0.0108	20.7 (<.001)	85.5	
Artificial neural network	4	1686	0.66-0.99	0.811 (0.63-0.94)	0.0436	58.3 (<.001)	94.9	
Random forest	3	1215	0.71-0.98	0.829 (0.61-0.97)	0.0456	43.1 (<.001)	95.4	
Decision tree	2	494	0.64-0.86	0.762 (0.52-0.94)	0.0319	29.6 (<.001)	96.6	
Naïve Bayes	2	990	0.68-0.97	0.841 (0.50-1.00)	0.0672	16.9 (<.001)	94.1	
<b>Number of stress classes</b>								.02 <sup>b</sup>
2	25	9222	0.58-0.97	0.773 (0.61-0.87)	0.0873	173.4 (<.001)	97.2	
>2	12	2203	0.86-0.99	0.944 (0.77-0.99)	0.2110	302.8 (<.001)	98.2	
<b>Type of WD<sup>c</sup></b>								.049
Smartwatches	11	2094	0.66-0.97	0.836 (0.53-0.95)	0.1466	91.6 (<.001)	96.6	
Smart bands	14	7128	0.58-0.77	0.695 (0.63-0.75)	0.0050	69.2 (<.001)	85.4	
Electrodes	12	2203	0.86-0.99	0.944 (0.77-0.99)	0.2110	302.8 (<.001)	98.2	
<b>Location of WD</b>								.02
Wrist	25	9222	0.58-0.97	0.773 (0.61-0.87)	0.0873	173.4 (<.001)	97.2	
Nonwrist	12	2203	0.86-0.99	0.944 (0.77-0.99)	0.2110	302.8 (<.001)	98.2	
<b>Data set size</b>								.009
≤100	14	789	0.80-0.99	0.947 (0.81-0.99)	0.1704	166.7 (<.001)	94.1	
>100	23	10,627	0.58-0.91	0.769 (0.62-0.86)	0.0758	485.3 (<.001)	97.4	
<b>Data source</b>								.57
WD-based	12	1825	0.80-0.97	0.895 (0.87-0.91)	0.0000	29.9 (<.001)	60.5	
WD-based data+other data	25	9600	0.58-0.99	0.830 (0.50-0.95)	0.3708	968.9 (<.001)	99.4	
<b>Data type</b>								.21
Heart rate data	13	5970	0.65-0.97	0.832 (0.52-0.95)	0.1594	115.4 (<.001)	98.6	

Group	Number of studies <sup>a</sup>	Sample size, n	Accuracy (%), range	Pooled mean accuracy (%), mean (95% CI)	Heterogeneity measures			Test for subgroup differences ( <i>P</i> value)
					$\tau^2$	Q ( <i>P</i> value)	<i>I</i> <sup>2</sup> (%)	
Heart rate data+EDA <sup>d</sup> data	12	2203	0.86-0.99	0.944 (0.77-0.99)	0.2110	302.8 (<.001)	98.2	
<b>Stress inducers</b>								.10
Yes	25	8173	0.65-0.99	0.902 (0.75-0.96)	0.2292	1137.8 (<.001)	98.9	
No	12	3252	0.58-0.80	0.697 (0.62-0.76)	0.0060	51.4 (<.001)	80.6	
<b>Ground truth</b>								.001
Objective	19	2413	0.80-0.99	0.933 (0.84-0.97)	0.1266	320.2 (<.001)	96.3	
Subjective	18	9012	0.58-0.80	0.706 (0.66-0.75)	0.0031	95.0 (<.001)	85.1	
<b>Validation method</b>								.16
K-fold	16	4087	0.66-0.99	0.904 (0.64-0.98)	0.3324	784.7 (<.001)	99.0	
Training-test split	14	7128	0.58-0.77	0.695 (0.63-0.75)	0.0050	69.2 (<.001)	85.4	
All studies	37	11,425	0.58-0.99	0.856 (0.698-0.934)	0.2402	1318.2 (<.001)	98.8	N/A <sup>e</sup>

<sup>a</sup>Many studies were included more than once in all meta-analyses except for the meta-analysis related to algorithms given that the studies assessed the performance of more than one algorithm.

<sup>b</sup>Italics indicate statistical significance (*P*<.05).

<sup>c</sup>WD: wearable device.

<sup>d</sup>EDA: electrodermal activity.

<sup>e</sup>N/A: not applicable.

### Other Measures

The sensitivity of wearable AI in detecting stress was reported in 32% (6/19) of the studies. From these studies, 27 sensitivity estimates were identified given that we extracted the highest sensitivity for each algorithm in each study. The estimates of sensitivity varied between 0.42 and 0.97, with an average of 0.755 (SD 0.181). From the 11% (2/19) of studies that reported specificity, 8 estimates of specificity were extracted given that more than one algorithm was assessed in these studies. The estimates of specificity ranged between 0.475 and 0.95, with an average of 0.744 (SD 0.147). We identified 46 estimates of *F*<sub>1</sub>-score from 53% (10/19) of the studies. The estimates of the *F*<sub>1</sub>-score ranged from 0.514 to 0.979, with an average of 0.759 (SD 0.139).

## Discussion

### Principal Findings

This review assessed the performance of wearable AI in detecting stress among students. Our analyses showed that wearable AI has an acceptable performance in detecting stress among students, but there is still room for improvement.

Specifically, our meta-analyses revealed that wearable AI was able to correctly classify students with and without stress in 85.6% of cases. Furthermore, our narrative synthesis showed that the performance of wearable AI in detecting students with stress (75.5%) is comparable with its performance in detecting students without stress (74.4%). We considered that the performance of wearable AI in detecting stress among students is suboptimal given that many previous reviews have shown a higher performance of AI in detecting other diseases or disorders, such as cancers [65-68], heart diseases [69,70], ear diseases [71], and ophthalmic disorders [72,73]. Moreover, relying solely on pooled mean accuracy is insufficient for drawing definitive conclusions regarding the performance of wearable AI; thus, we considered sensitivity and specificity when evaluating its performance, which were lower than the pooled accuracy.

The subgroup analyses in this review demonstrated that the performance of wearable AI is moderated by 5 factors. The first factor is the type of wearable device used to collect digital biomarkers. To be more precise, electrodes have higher accuracy (94.4%) in detecting stress than smartwatches (83.6%) and smart bands (69.5%). This can be explained by the practice of typically positioning electrodes on the most suitable areas of the body

for gathering specific biomarkers, thereby collecting more accurate data, as opposed to most commercial and fashionable popular devices being restricted to wrists.

The second moderator identified in this review is the placement of wearable devices on the body. To elaborate, studies that used wrist-worn devices exhibited lower accuracy in detecting stress among students compared with those that used wearable devices placed on other parts of the body (94.4% vs 77.3%). This can be attributed to 4 factors. First, all studies that used non-wrist-worn devices collected data from more than one area of the body (eg, chest, fingers, and palm) and used electrodes, whereas all studies that used wrist-worn devices collected data from only one area (ie, wrist) and used smartwatches or smart bands. Second, the placement of wearable devices on the wrist may lead to less accurate data collection because of the frequent arm movements. Third, all studies that used non-wrist-worn devices collected EDA data, whereas none of the studies that used wrist-worn devices collected EDA data. Finally, the ground truth in all studies that used non-wrist-worn devices was assessed objectively, whereas the ground truth in approximately half of the studies that used wrist-worn devices was identified subjectively.

The third moderator identified in this review is the number of stress classes that wearable AI endeavors to detect. Surprisingly, our analysis revealed that wearable AI exhibits greater accuracy in classifying more than 2 stress levels (94.4%) as opposed to just 2 stress classes (77.3%). This may be attributed to the same 4 aforementioned factors given that the estimates used in the subgroup meta-analysis to assess the performance of wearable AI in classifying more than 2 stress levels were the exact estimates used in the meta-analysis to evaluate the performance of non-wrist-worn devices.

The fourth moderator found in this review is the ground truth. Specifically, the efficacy of wearable AI in stress detection tends to be superior when the ground truth is determined through objective assessments (eg, the experimental context), in contrast to when it relies on subjective assessments (ie, self-reported questionnaires; 93.3% vs 70.6%). This is expected as objective assessments are generally more reliable and accurate than subjective assessments.

The last moderator identified in this review is the data set size. Unexpectedly, our subgroup meta-analyses demonstrated that the accuracy of wearable AI in detecting stress is higher when using a data set size of  $\leq 100$  (94.7%) as opposed to data sets with sizes of  $> 100$  (76.9%). Although this seems odd at first glance, the reason could be that the ground truth in all studies that used a data set size of  $\leq 100$  was examined objectively, whereas the ground truth in approximately half of the studies that used a data set size of  $> 100$  was identified subjectively.

As mentioned in the *Introduction* section, none of the previous reviews statistically synthesized the results of previous studies. Thus, we calculated the traditional average of accuracy estimates reported in the studies included in these reviews. We noticed comparable findings despite the differences between our review and these reviews. To elaborate, although the pooled mean in this review was 85.6%, the average accuracy estimates in the previous reviews were 87.6% (50%-100%) [28], 87%

(70.8%-99%) [34], 86% (60%-100%) [15], 85.4% (64.5%-98.3%) [35], 84.2% (51.2%-100%) [24], and 82.9% (53%-99%) [13]. Furthermore, this review showed findings comparable with those of previous studies on the performance of wearable AI in detecting depression [74] and anxiety [75]. Specifically, the pooled mean accuracies of wearable AI in detecting depression and anxiety were 89% [74] and 82% [75], respectively.

## Research and Practical Implications

This review found that wearable AI holds promise as a valuable tool for detecting stress among students. Nonetheless, we are unable to endorse the immediate integration of wearable AI into clinical and academic practices for the following reasons: (1) its efficiency in detecting stress among students is currently suboptimal, (2) the number of participants was small ( $\leq 50$ ) in 74% (14/19) of the studies, (3) the accuracy estimates included in all meta-analyses were extracted from a few studies ( $\leq 6$ ), and (4) only 21% (4/19) of the studies were judged to have a low risk of bias in all domains. Hence, it is advisable to use wearable AI in conjunction with other clinical assessments such as self-report questionnaires to detect stress among students.

Most of the studies included in this review (17/19, 89%) targeted undergraduate students, although students at other educational levels are vulnerable to stress. Hence, future research endeavors should recruit students at different educational levels, such as postgraduate and high school students. This review primarily focused on stress among students even though there is a substantial body of research on the effectiveness of wearable AI in identifying stress in other demographic groups. Thus, further systematic reviews are needed to consolidate and synthesize the evidence on the performance of wearable AI in detecting stress across diverse populations.

In this review, all the included studies evaluated the effectiveness of wearable AI in detecting current stress levels rather than predicting the occurrence of stress in the future. Anticipating the onset of stress in the future is equally if not more crucial than identifying the current stress state as this can facilitate the development and implementation of more efficient, timely, and tailored interventions. For this reason, we urge researchers to conduct further studies on the potential of wearable AI for predicting future occurrences of stress.

None of the studies included in this review evaluated the effectiveness of wearable AI in differentiating stress from other mental health conditions such as depression and anxiety, nor did they assess its ability to distinguish between different types of stress (eg, acute stress, chronic stress, emotional stress, physical stress, eustress, distress, and posttraumatic stress disorder). In clinical practice, health care professionals often rely on complex and error-prone diagnostic methods to make distinctions among various patient groups rather than merely distinguishing them from individuals without health issues. Consequently, it is imperative for researchers to investigate how well wearable AI can differentiate between different types of stress and distinguish individuals with stress from those who have other mental disorders exhibiting signs and symptoms similar to those of stress.

Although previous studies have examined the performance of wearable AI based on different factors (eg, algorithms, data types, and number of stress classes), they did not assess its performance based on the placement of the wearable device (eg, wrist, chest, or fingers), the methods used to induce stress (eg, Stroop test, arithmetic tasks, and physical exercises), or the methods used to assess the ground truth (eg, subjective and objective assessments). We strongly encourage researchers to consider these factors in future studies. Furthermore, none of the previous studies have used biological samples as the ground truth, although they are considered as the gold-standard test of stress. For example, elevated cortisol levels in blood, saliva, or urine samples can be indicative of chronic stress. Such gold-standard tests, especially those that can be measured easily and noninvasively, should be considered in future studies.

Even though this review included 19 studies, we were able to include results from only 6 (32%) in our meta-analyses as the remaining studies did not provide sufficient details essential for conducting meta-analyses (eg, confusion matrices, number of cases, and number of controls). In addition, these studies did not present multiple performance measures such as accuracy, sensitivity, specificity, and  $F_1$ -score, which are necessary for us to estimate the required details. Hence, researchers should include the aforementioned details in their reports to facilitate the conduct of meta-analyses by others.

A total of 74% (14/19) of the studies had a limited number of participants, with  $\leq 50$  individuals involved. This could have posed challenges in identifying potential performance differences in wearable AI when conducting subgroup meta-analyses. In addition, this could have acted as a hindrance for the development of algorithms that rely on a substantial amount of data. To address these issues, it is essential for researchers to recruit a larger number of participants, ensuring robust statistical power and enabling the use of more advanced and efficient algorithms.

### Limitations

This review excluded studies that assessed the performance of (1) nonwearable devices, handheld devices, near-body wearable

devices, in-body wearable devices, wearable devices wired to nonwearable devices, and wearables requiring expert oversight; (2) wearable AI in detecting other mental disorders (eg, attention-deficit/hyperactivity disorder, bipolar disorder, and schizophrenia); (3) wearable AI in detecting stress among populations other than students; and (4) wearable AI in predicting outcomes of stress treatment. Consequently, the generalizability of our findings to such devices, disorders, populations, and outcomes is limited.

This review could not draw definitive conclusions based on the results of the meta-analyses for several reasons. First, it is probable that we overlooked some studies as we excluded studies that were published in non-English languages and before 2015 and used publicly available data sets (eg, WESAD) or data sets from previous studies. Second, the accuracy estimates included in all meta-analyses were drawn from a limited number of studies ( $\leq 6$ ), primarily as the remaining studies lacked the necessary information required for conducting meta-analyses.

### Conclusions

Wearable AI holds promise as a valuable tool for detecting stress among students. However, it is not sufficiently ready to be integrated into clinical and academic practices given its suboptimal performance. Until further evidence demonstrates an ideal performance of wearable AI, it should be used in conjunction with other clinical assessments to help detect stress among students. Given the limitations of the current evidence, researchers should investigate how well wearable AI can differentiate between different types of stress and distinguish individuals with stress from those who have other mental disorders exhibiting signs and symptoms similar to those of stress. Furthermore, future research should recruit students at different educational levels, such as postgraduate and high school students. There is also a need to conduct studies on the potential of wearable AI for predicting future occurrences of stress. Future studies should compare the performance of wearable AI based on the placement of the wearable device, the methods used to induce stress, and the methods used to assess the ground truth. Finally, researchers should provide detailed results to facilitate the conduct of meta-analyses by others.

---

### Acknowledgments

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

---

### Data Availability

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

---

### Authors' Contributions

A Abd-alrazaq, A Ahmed, R AlSaad, and SA developed the protocol with guidance from and under the supervision of JS. A Abd-alrazaq searched the electronic databases and conducted backward and forward reference list checking. The study selection process, data extraction, and risk-of-bias assessment were carried out by MA and R Ahmad under the supervision of R AlSaad and A Abd-alrazaq. Data synthesis was conducted by A Abd-alrazaq. The *Introduction* section was written by MA and RD. The *Methods* section was written by A Ahmed. The *Results* section was written by A Abd-alrazaq and R AlSaad. The *Discussion* and *Conclusions* sections were written by A Abd-alrazaq. The manuscript was critically revised 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.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

PRISMA-DTA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Diagnostic Test Accuracy) checklist.

[\[DOC File , 68 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Search strategy.

[\[DOCX File , 42 KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

Data extraction form.

[\[DOCX File , 22 KB-Multimedia Appendix 3\]](#)

## Multimedia Appendix 4

Modified version of the Quality Assessment of Diagnostic Accuracy Studies–Revised.

[\[DOCX File , 26 KB-Multimedia Appendix 4\]](#)

## Multimedia Appendix 5

Characteristics of each included study.

[\[DOCX File , 35 KB-Multimedia Appendix 5\]](#)

## Multimedia Appendix 6

Features of wearable devices.

[\[DOCX File , 34 KB-Multimedia Appendix 6\]](#)

## Multimedia Appendix 7

Features of the artificial intelligence algorithms.

[\[DOCX File , 38 KB-Multimedia Appendix 7\]](#)

## Multimedia Appendix 8

Reviewers' judgments about each risk of bias and applicability domain for each included study.

[\[DOCX File , 42 KB-Multimedia Appendix 8\]](#)

## References

1. Liu Z, Xie Y, Sun Z, Liu D, Yin H, Shi L. Factors associated with academic burnout and its prevalence among university students: a cross-sectional study. *BMC Med Educ*. May 06, 2023;23(1):317. [\[FREE Full text\]](#) [doi: [10.1186/s12909-023-04316-y](https://doi.org/10.1186/s12909-023-04316-y)] [Medline: [37149602](https://pubmed.ncbi.nlm.nih.gov/37149602/)]
2. Yang G, Wang Z, Wu W. Social comparison orientation and mental health: the mediating role of psychological capital. *soc behav pers*. Jan 06, 2021;49(1):1-11. [\[FREE Full text\]](#) [doi: [10.2224/sbp.9767](https://doi.org/10.2224/sbp.9767)]
3. Kumaraswamy N. Academic stress, anxiety and depression among college students: a brief review. *Int Rev Soc Sci humanit*. 2013;5(1):135-143.
4. Rith-Najarian LR, Boustani MM, Chorpita BF. A systematic review of prevention programs targeting depression, anxiety, and stress in university students. *J Affect Disord*. Oct 01, 2019;257:568-584. [doi: [10.1016/j.jad.2019.06.035](https://doi.org/10.1016/j.jad.2019.06.035)] [Medline: [31326690](https://pubmed.ncbi.nlm.nih.gov/31326690/)]
5. Mofatteh M. Risk factors associated with stress, anxiety, and depression among university undergraduate students. *AIMS Public Health*. 2021;8(1):36-65. [\[FREE Full text\]](#) [doi: [10.3934/publichealth.2021004](https://doi.org/10.3934/publichealth.2021004)] [Medline: [33575406](https://pubmed.ncbi.nlm.nih.gov/33575406/)]
6. Bedewy D, Gabriel A. Examining perceptions of academic stress and its sources among university students: the perception of academic stress scale. *Health Psychol Open*. Jul 30, 2015;2(2):2055102915596714. [\[FREE Full text\]](#) [doi: [10.1177/2055102915596714](https://doi.org/10.1177/2055102915596714)] [Medline: [28070363](https://pubmed.ncbi.nlm.nih.gov/28070363/)]



7. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. *J Affect Disord*. May 15, 2013;148(1):1-11. [doi: [10.1016/j.jad.2012.11.026](https://doi.org/10.1016/j.jad.2012.11.026)] [Medline: [23246209](https://pubmed.ncbi.nlm.nih.gov/23246209/)]
8. Shiralkar MT, Harris TB, Eddins-Folensbee FF, Coverdale JH. A systematic review of stress-management programs for medical students. *Acad Psychiatry*. May 01, 2013;37(3):158-164. [doi: [10.1176/appi.ap.12010003](https://doi.org/10.1176/appi.ap.12010003)] [Medline: [23446664](https://pubmed.ncbi.nlm.nih.gov/23446664/)]
9. Hobfoll SE. *Stress, Culture, and Community: The Psychology and Philosophy of Stress*. New York, NY. Springer; 2004.
10. Lazarus RS. From psychological stress to the emotions: a history of changing outlooks. *Annu Rev Psychol*. Jan 1993;44(1):1-21. [doi: [10.1146/annurev.ps.44.020193.000245](https://doi.org/10.1146/annurev.ps.44.020193.000245)] [Medline: [8434890](https://pubmed.ncbi.nlm.nih.gov/8434890/)]
11. Schneiderman N, Ironson G, Siegel SD. Stress and health: psychological, behavioral, and biological determinants. *Annu Rev Clin Psychol*. Apr 01, 2005;1(1):607-628. [FREE Full text] [doi: [10.1146/annurev.clinpsy.1.102803.144141](https://doi.org/10.1146/annurev.clinpsy.1.102803.144141)] [Medline: [17716101](https://pubmed.ncbi.nlm.nih.gov/17716101/)]
12. Edjah K, Ankomah F, Domey E, Laryea JE. Stress and its impact on academic and social life of undergraduate university students in Ghana: a structural equation modeling approach. *Open Educ Stud*. Jun 2020;2(1):37-44. [FREE Full text] [doi: [10.1515/edu-2020-0100](https://doi.org/10.1515/edu-2020-0100)]
13. Can YS, Arnrich B, Ersoy C. Stress detection in daily life scenarios using smart phones and wearable sensors: a survey. *J Biomed Inform*. Apr 2019;92:103139. [FREE Full text] [doi: [10.1016/j.jbi.2019.103139](https://doi.org/10.1016/j.jbi.2019.103139)] [Medline: [30825538](https://pubmed.ncbi.nlm.nih.gov/30825538/)]
14. Aldwin CM. *Stress, Coping, and Development: An Integrative Perspective*. New York, NY. The Guilford Press; 2009.
15. Elzeiny S, Qaraq M. Machine learning approaches to automatic stress detection: a review. In: *Proceedings of the IEEE/ACS 15th International Conference on Computer Systems and Applications*. Presented at: AICCSA '18; October 28-November 1, 2018, 2018;1-6; Aqaba, Jordan. URL: <https://ieeexplore.ieee.org/document/8612825> [doi: [10.1109/aiccsa.2018.8612825](https://doi.org/10.1109/aiccsa.2018.8612825)]
16. Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. *Sports Med*. Jan 13, 2014;44(1):81-121. [FREE Full text] [doi: [10.1007/s40279-013-0090-5](https://doi.org/10.1007/s40279-013-0090-5)] [Medline: [24030837](https://pubmed.ncbi.nlm.nih.gov/24030837/)]
17. Farrell AK, Simpson JA. Effects of relationship functioning on the biological experience of stress and physical health. *Curr Opin Psychol*. Feb 2017;13:49-53. [doi: [10.1016/j.copsyc.2016.04.014](https://doi.org/10.1016/j.copsyc.2016.04.014)] [Medline: [28813293](https://pubmed.ncbi.nlm.nih.gov/28813293/)]
18. Sandi C. Stress, cognitive impairment and cell adhesion molecules. *Nat Rev Neurosci*. Dec 2004;5(12):917-930. [doi: [10.1038/nrn1555](https://doi.org/10.1038/nrn1555)] [Medline: [15550947](https://pubmed.ncbi.nlm.nih.gov/15550947/)]
19. Jonsdottir IH, Nordlund A, Ellbin S, Ljung T, Glise K, Währborg P, et al. Cognitive impairment in patients with stress-related exhaustion. *Stress*. Mar 17, 2013;16(2):181-190. [doi: [10.3109/10253890.2012.708950](https://doi.org/10.3109/10253890.2012.708950)] [Medline: [22746338](https://pubmed.ncbi.nlm.nih.gov/22746338/)]
20. Sandi C, Haller J. Stress and the social brain: behavioural effects and neurobiological mechanisms. *Nat Rev Neurosci*. May 20, 2015;16(5):290-304. [FREE Full text] [doi: [10.1038/nrn3918](https://doi.org/10.1038/nrn3918)] [Medline: [25891510](https://pubmed.ncbi.nlm.nih.gov/25891510/)]
21. Gradus JL. Prevalence and prognosis of stress disorders: a review of the epidemiologic literature. *Clin Epidemiol*. May 03, 2017;9:251-260. [FREE Full text] [doi: [10.2147/CLEP.S106250](https://doi.org/10.2147/CLEP.S106250)] [Medline: [28496365](https://pubmed.ncbi.nlm.nih.gov/28496365/)]
22. Vrshek-Schallhorn S, Stroud CB, Mineka S, Hammen C, Zinbarg RE, Wolitzky-Taylor K, et al. Chronic and episodic interpersonal stress as statistically unique predictors of depression in two samples of emerging adults. *J Abnorm Psychol*. Nov 2015;124(4):918-932. [FREE Full text] [doi: [10.1037/abn0000088](https://doi.org/10.1037/abn0000088)] [Medline: [26301973](https://pubmed.ncbi.nlm.nih.gov/26301973/)]
23. Kumar S, Bhukar JP. Stress level and coping strategies of college students. *J Physic Educ Sport Manag*. Jan 2013;4(1):5-11. [FREE Full text]
24. Arsalan A, Anwar S, Majid M. Mental stress detection using data from wearable and non-wearable sensors: a review. arXiv. Preprint posted online on Feb 7, 2022. [FREE Full text]
25. Frisone F, Sicari F, Settineri S, Merlo EM. Clinical psychological assessment of stress: a narrative review of the last 5 years. *Clin Neuropsychiatry*. Apr 2021;18(2):91-100. [FREE Full text] [doi: [10.36131/cnfioritiditore20210203](https://doi.org/10.36131/cnfioritiditore20210203)] [Medline: [34909024](https://pubmed.ncbi.nlm.nih.gov/34909024/)]
26. Epel ES, Crosswell AD, Mayer SE, Prather AA, Slavich GM, Puterman E, et al. More than a feeling: a unified view of stress measurement for population science. *Front Neuroendocrinol*. Apr 2018;49:146-169. [FREE Full text] [doi: [10.1016/j.yfrne.2018.03.001](https://doi.org/10.1016/j.yfrne.2018.03.001)] [Medline: [29551356](https://pubmed.ncbi.nlm.nih.gov/29551356/)]
27. Dhama K, Latheef SK, Dadar M, Samad HA, Munjal A, Khandia R, et al. Biomarkers in stress related diseases/disorders: diagnostic, prognostic, and therapeutic values. *Front Mol Biosci*. Oct 18, 2019;6:91. [FREE Full text] [doi: [10.3389/fmolb.2019.00091](https://doi.org/10.3389/fmolb.2019.00091)] [Medline: [31750312](https://pubmed.ncbi.nlm.nih.gov/31750312/)]
28. Gedam S, Paul S. A review on mental stress detection using wearable sensors and machine learning techniques. *IEEE Access*. 2021;9:84045-84066. [FREE Full text] [doi: [10.1109/access.2021.3085502](https://doi.org/10.1109/access.2021.3085502)]
29. Ahmed A, Aziz S, Abd-Alrazaq A, Farooq F, Sheikh J. Overview of artificial intelligence-driven wearable devices for diabetes: scoping review. *J Med Internet Res*. Aug 09, 2022;24(8):e36010. [FREE Full text] [doi: [10.2196/36010](https://doi.org/10.2196/36010)] [Medline: [35943772](https://pubmed.ncbi.nlm.nih.gov/35943772/)]
30. Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, et al. Wearable artificial intelligence for anxiety and depression: scoping review. *J Med Internet Res*. Jan 19, 2023;25:e42672. [FREE Full text] [doi: [10.2196/42672](https://doi.org/10.2196/42672)] [Medline: [36656625](https://pubmed.ncbi.nlm.nih.gov/36656625/)]
31. Iqbal SM, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. *Npj Flex Electron*. Apr 12, 2021;5(1):9. [FREE Full text] [doi: [10.1038/s41528-021-00107-x](https://doi.org/10.1038/s41528-021-00107-x)]

32. Hickey BA, Chalmers T, Newton P, Lin CT, Sibbritt D, McLachlan CS, et al. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: a systematic review. *Sensors (Basel)*. May 16, 2021;21(10):3461. [FREE Full text] [doi: [10.3390/s21103461](https://doi.org/10.3390/s21103461)] [Medline: [34065620](https://pubmed.ncbi.nlm.nih.gov/34065620/)]
33. Long N, Lei Y, Peng L, Xu P, Mao P. A scoping review on monitoring mental health using smart wearable devices. *Math Biosci Eng*. May 27, 2022;19(8):7899-7919. [FREE Full text] [doi: [10.3934/mbe.2022369](https://doi.org/10.3934/mbe.2022369)] [Medline: [35801449](https://pubmed.ncbi.nlm.nih.gov/35801449/)]
34. Gradl S, Wirth M, Richer R, Rohleder N, Eskofier B. An overview of the feasibility of permanent, real-time, unobtrusive stress measurement with current wearables. In: *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. Presented at: PervasiveHealth '19; May 20-23, 2019, 2019;360-365; Trento, Italy. URL: <https://dl.acm.org/doi/10.1145/3329189.3329233> [doi: [10.1145/3329189.3329233](https://doi.org/10.1145/3329189.3329233)]
35. Vos G, Trinh K, Sarnyai Z, Rahimi Azghadi M. Generalizable machine learning for stress monitoring from wearable devices: a systematic literature review. *Int J Med Inform*. May 2023;173:105026. [FREE Full text] [doi: [10.1016/j.ijmedinf.2023.105026](https://doi.org/10.1016/j.ijmedinf.2023.105026)] [Medline: [36893657](https://pubmed.ncbi.nlm.nih.gov/36893657/)]
36. McInnes MD, Moher D, Thombs BD, McGrath TA, Bossuyt PM, the PRISMA-DTA Group; et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. Jan 23, 2018;319(4):388-396. [doi: [10.1001/jama.2017.19163](https://doi.org/10.1001/jama.2017.19163)] [Medline: [29362800](https://pubmed.ncbi.nlm.nih.gov/29362800/)]
37. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. Oct 18, 2011;155(8):529-536. [FREE Full text] [doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)] [Medline: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)]
38. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST Group. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. Jan 01, 2019;170(1):51-58. [FREE Full text] [doi: [10.7326/M18-1376](https://doi.org/10.7326/M18-1376)] [Medline: [30596875](https://pubmed.ncbi.nlm.nih.gov/30596875/)]
39. Moltu C, Stefansen J, Svisdahl M, Veseth M. *Doing Meta-Analysis With R: A Hands-On Guide*. Boca Raton, FL. CRC Press; 2012.
40. Borenstein M, Hedges LV, Higgins PT, Rothstein HR. *Introduction to Meta - Analysis*. Hoboken, NJ. John Wiley & Sons; 2009.
41. Freeman MF, Tukey JW. Transformations related to the angular and the square root. *Ann Math Statist*. Dec 1950;21(4):607-611. [FREE Full text] [doi: [10.1214/aoms/1177729756](https://doi.org/10.1214/aoms/1177729756)]
42. Schwarzer G. Meta: an R package for meta-analysis. *R news*. 2007;7(3):40-45. [FREE Full text]
43. Assink M, Wibbelink CJ. Fitting three-level meta-analytic models in R: a step-by-step tutorial. *Quant Method Psychol*. Oct 01, 2016;12(3):154-174. [FREE Full text] [doi: [10.20982/tqmp.12.3.p154](https://doi.org/10.20982/tqmp.12.3.p154)]
44. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. Sep 06, 2003;327(7414):557-560. [FREE Full text] [doi: [10.1136/bmj.327.7414.557](https://doi.org/10.1136/bmj.327.7414.557)] [Medline: [12958120](https://pubmed.ncbi.nlm.nih.gov/12958120/)]
45. Deeks J, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Hoboken, NJ. John Wiley & Sons; 2019;241-284.
46. Altaf H, Ibrahim SN, Olanrewaju RF. Non invasive stress detection method based on discrete wavelet transform and machine learning algorithms. In: *Proceedings of the 11th IEEE Symposium on Computer Applications & Industrial Electronics*. Presented at: ISCAIE '21; April 3-4, 2021, 2021;106-111; Penang, Malaysia. URL: <https://ieeexplore.ieee.org/document/9431805> [doi: [10.1109/iscaie51753.2021.9431805](https://doi.org/10.1109/iscaie51753.2021.9431805)]
47. Aqajari SA, Labbaf S, Tran PH, Nguyen B, Mehrabadi MA, Levorato M, et al. Context-aware stress monitoring using wearable and mobile technologies in everyday settings. medRxiv. Preprint posted online on Apr 24, 2023. [FREE Full text] [doi: [10.1101/2023.04.20.23288181](https://doi.org/10.1101/2023.04.20.23288181)]
48. Can YS, Chalabianloo N, Ekiz D, Ersoy C. Continuous stress detection using wearable sensors in real life: algorithmic programming contest case study. *Sensors (Basel)*. Apr 18, 2019;19(8):1849. [FREE Full text] [doi: [10.3390/s19081849](https://doi.org/10.3390/s19081849)] [Medline: [31003456](https://pubmed.ncbi.nlm.nih.gov/31003456/)]
49. Coutts LV, Plans D, Brown AW, Collomosse J. Deep learning with wearable based heart rate variability for prediction of mental and general health. *J Biomed Inform*. Dec 2020;112:103610. [FREE Full text] [doi: [10.1016/j.jbi.2020.103610](https://doi.org/10.1016/j.jbi.2020.103610)] [Medline: [33137470](https://pubmed.ncbi.nlm.nih.gov/33137470/)]
50. de Arriba Pérez F, Santos-Gago JM, Caeiro-Rodríguez M, Fernández Iglesias MJ. Evaluation of commercial-off-the-shelf wrist wearables to estimate stress on students. *J Vis Exp*. Jun 16, 2018(136):57590. [FREE Full text] [doi: [10.3791/57590](https://doi.org/10.3791/57590)] [Medline: [29985338](https://pubmed.ncbi.nlm.nih.gov/29985338/)]
51. Egilmez B, Poyraz E, Zhou W, Memik G, Dinda P, Alshurafa N. UStress: understanding college student subjective stress using wrist-based passive sensing. In: *Proceedings of the 2021 International Conference on Pervasive Computing and Communications Workshops*. Presented at: PerCom '21; March 22-26, 2021, 2021;673-678; Kona, HI. URL: <https://ieeexplore.ieee.org/document/7917644> [doi: [10.1109/percomw.2017.7917644](https://doi.org/10.1109/percomw.2017.7917644)]
52. Hasanbasic A, Spahic M, Bosnjic D, Adzic HH, Mesic V, Jahic O. Recognition of stress levels among students with wearable sensors. In: *Proceedings of the 18th International Symposium on INFOTEH-JAHORINA*. Presented at: INFOTEH '19; March 20-22, 2019, 2019;1-4; East Sarajevo, Bosnia and Herzegovina. URL: <https://ieeexplore.ieee.org/document/8717754> [doi: [10.1109/infotech.2019.8717754](https://doi.org/10.1109/infotech.2019.8717754)]

53. Islam TZ, Liang PW, Sweeney F, Pragner C, Thiagarajan JJ, Sharmin M, et al. College life is hard! - shedding light on stress prediction for autistic college students using data-driven analysis. In: Proceedings of the 45th Annual Computers, Software, and Applications Conference. Presented at: COMPSAC '21; July 12-16, 2021, 2021;428-437; Madrid, Spain. URL: <https://ieeexplore.ieee.org/document/9529720> [doi: [10.1109/compsac51774.2021.00066](https://doi.org/10.1109/compsac51774.2021.00066)]
54. Jafarlou S, Lai J, Azimi I, Mousavi Z, Labbaf S, Jain RC, et al. Objective prediction of next-day's affect using multimodal physiological and behavioral data: algorithm development and validation study. *JMIR Form Res*. Mar 15, 2023;7:e39425. [FREE Full text] [doi: [10.2196/39425](https://doi.org/10.2196/39425)] [Medline: [36920456](https://pubmed.ncbi.nlm.nih.gov/36920456/)]
55. Jiang JY, Chao Z, Bertozzi AL, Wang W, Young SD, Needell D. Learning to predict human stress level with incomplete sensor data from wearable devices. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Presented at: CIKM '19; November 3-7, 2019, 2019;2773-2781; Beijing, China. URL: <https://dl.acm.org/doi/10.1145/3357384.3357831> [doi: [10.1145/3357384.3357831](https://doi.org/10.1145/3357384.3357831)]
56. Moridani MK, Mahabadi Z, Javadi N. Heart rate variability features for different stress classification. *Bratisl Lek Listy*. 2020;121(9):619-627. [doi: [10.4149/BLL\\_2020\\_107](https://doi.org/10.4149/BLL_2020_107)] [Medline: [32990009](https://pubmed.ncbi.nlm.nih.gov/32990009/)]
57. Nagar P, Sethia D. Brain mapping based stress identification using portable EEG based device. In: Proceedings of the 11th International Conference on Communication Systems & Networks. Presented at: COMSNETS '19; April 23-24, 2019, 2019;601-606; Bengaluru, India. URL: <https://ieeexplore.ieee.org/document/8711009> [doi: [10.1109/comsnets.2019.8711009](https://doi.org/10.1109/comsnets.2019.8711009)]
58. Padmaja B, Prasad VV, Sunitha KV, Reddy CS, Anil CH. DetectStress: a novel stress detection system based on smartphone and wireless physical activity tracker. In: Proceedings of the 1st International Conference on Artificial Intelligence and Cognitive Computing. Presented at: AICC '18; November 24-25, 2018, 2018;67-80; Virtual Event. [doi: [10.1007/978-981-13-1580-0\\_7](https://doi.org/10.1007/978-981-13-1580-0_7)]
59. Sandulescu V. Stress detection using wearable physiological sensors. In: Proceedings of the 2015 International Work-Conference on the Interplay Between Natural and Artificial Computation on Artificial Computation in Biology and Medicine. Presented at: IWINAC '15; June 1-5, 2015, 2015;526-532; Elche, Spain. URL: <http://tinyurl.com/3xjmecjm> [doi: [10.1007/978-3-319-18914-7\\_55](https://doi.org/10.1007/978-3-319-18914-7_55)]
60. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. *J Med Internet Res*. Jun 08, 2018;20(6):e210. [FREE Full text] [doi: [10.2196/jmir.9410](https://doi.org/10.2196/jmir.9410)] [Medline: [29884610](https://pubmed.ncbi.nlm.nih.gov/29884610/)]
61. Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction. Presented at: ICMI '18; October 16-20, 2018, 2018;400-408; Boulder, CO. URL: <https://dl.acm.org/doi/10.1145/3242969.3242985> [doi: [10.1145/3242969.3242985](https://doi.org/10.1145/3242969.3242985)]
62. Silva E, Aguiar J, Reis LP, Sá JO, Gonçalves J, Carvalho V. Stress among Portuguese medical students: the EuStress solution. *J Med Syst*. Jan 02, 2020;44(2):45. [doi: [10.1007/s10916-019-1520-1](https://doi.org/10.1007/s10916-019-1520-1)] [Medline: [31897774](https://pubmed.ncbi.nlm.nih.gov/31897774/)]
63. Tiwari S, Agarwal S. A shrewd artificial neural network-based hybrid model for pervasive stress detection of students using galvanic skin response and electrocardiogram signals. *Big Data*. Dec 01, 2021;9(6):427-442. [doi: [10.1089/big.2020.0256](https://doi.org/10.1089/big.2020.0256)]
64. Wu Y, Daoudi M, Amad A, Sparrow L, D'Hondt F. Unsupervised learning method for exploring students' mental stress in medical simulation training. In: Companion Publication of the 2020 International Conference on Multimodal Interaction. Presented at: ICMI '20; October 25-29, 2020, 2020;165-170; Virtual Event. URL: <https://dl.acm.org/doi/10.1145/3395035.3425191> [doi: [10.1145/3395035.3425191](https://doi.org/10.1145/3395035.3425191)]
65. Zhang SM, Wang YJ, Zhang ST. Accuracy of artificial intelligence-assisted detection of esophageal cancer and neoplasms on endoscopic images: a systematic review and meta-analysis. *J Dig Dis*. Jun 18, 2021;22(6):318-328. [FREE Full text] [doi: [10.1111/1751-2980.12992](https://doi.org/10.1111/1751-2980.12992)] [Medline: [33871932](https://pubmed.ncbi.nlm.nih.gov/33871932/)]
66. Jiang K, Jiang X, Pan J, Wen Y, Huang Y, Weng S, et al. Current evidence and future perspective of accuracy of artificial intelligence application for early gastric cancer diagnosis with endoscopy: a systematic and meta-analysis. *Front Med (Lausanne)*. Mar 15, 2021;8:629080. [FREE Full text] [doi: [10.3389/fmed.2021.629080](https://doi.org/10.3389/fmed.2021.629080)] [Medline: [33791323](https://pubmed.ncbi.nlm.nih.gov/33791323/)]
67. Dumitrescu EA, Ungureanu BS, Cazacu IM, Florescu LM, Streba L, Croitoru VM, et al. Diagnostic value of artificial intelligence-assisted endoscopic ultrasound for pancreatic cancer: a systematic review and meta-analysis. *Diagnostics (Basel)*. Jan 25, 2022;12(2):309. [FREE Full text] [doi: [10.3390/diagnostics12020309](https://doi.org/10.3390/diagnostics12020309)] [Medline: [35204400](https://pubmed.ncbi.nlm.nih.gov/35204400/)]
68. Morozov A, Taratkin M, Bazarkin A, Rivas JG, Puliatti S, Checcucci E, et al. Young Academic Urologists (YAU) Working Group in Uro-technology of the European Association of Urology. A systematic review and meta-analysis of artificial intelligence diagnostic accuracy in prostate cancer histology identification and grading. *Prostate Cancer Prostatic Dis*. Dec 25, 2023;26(4):681-692. [doi: [10.1038/s41391-023-00673-3](https://doi.org/10.1038/s41391-023-00673-3)] [Medline: [37185992](https://pubmed.ncbi.nlm.nih.gov/37185992/)]
69. Hoodbhoy Z, Jiwani U, Sattar S, Salam R, Hasan B, Das JK. Diagnostic accuracy of machine learning models to identify congenital heart disease: a meta-analysis. *Front Artif Intell*. Jul 8, 2021;4:708365. [FREE Full text] [doi: [10.3389/frai.2021.708365](https://doi.org/10.3389/frai.2021.708365)] [Medline: [34308341](https://pubmed.ncbi.nlm.nih.gov/34308341/)]
70. Li XM, Gao XY, Tse G, Hong SD, Chen KY, Li GP, et al. Electrocardiogram-based artificial intelligence for the diagnosis of heart failure: a systematic review and meta-analysis. *J Geriatr Cardiol*. Dec 28, 2022;19(12):970-980. [FREE Full text] [doi: [10.11909/j.issn.1671-5411.2022.12.002](https://doi.org/10.11909/j.issn.1671-5411.2022.12.002)] [Medline: [36632204](https://pubmed.ncbi.nlm.nih.gov/36632204/)]

71. Habib AR, Kajbafzadeh M, Hasan Z, Wong E, Gunasekera H, Perry C, et al. Artificial intelligence to classify ear disease from otoscopy: a systematic review and meta-analysis. *Clin Otolaryngol*. May 15, 2022;47(3):401-413. [FREE Full text] [doi: [10.1111/coa.13925](https://doi.org/10.1111/coa.13925)] [Medline: [35253378](https://pubmed.ncbi.nlm.nih.gov/35253378/)]
72. Bulloch G, Seth I, Zhu Z, Stapleton FJ, Fung J, Tan Z, et al. Artificial intelligence applications and performance metrics in ophthalmology: a systematic review and meta-analysis. *Vestn Oftalmol Vis Sci*. 2023;6(1):1037. [FREE Full text]
73. Wu JH, Nishida T, Weinreb RN, Lin JW. Performances of machine learning in detecting glaucoma using fundus and retinal optical coherence tomography images: a meta-analysis. *Am J Ophthalmol*. May 2022;237:1-12. [doi: [10.1016/j.ajo.2021.12.008](https://doi.org/10.1016/j.ajo.2021.12.008)] [Medline: [34942113](https://pubmed.ncbi.nlm.nih.gov/34942113/)]
74. Abd-Alrazaq A, AlSaad R, Shuweihdi F, Ahmed A, Aziz S, Sheikh J. Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. *NPJ Digit Med*. May 05, 2023;6(1):84. [FREE Full text] [doi: [10.1038/s41746-023-00828-5](https://doi.org/10.1038/s41746-023-00828-5)] [Medline: [37147384](https://pubmed.ncbi.nlm.nih.gov/37147384/)]
75. Abd-Alrazaq A, AlSaad R, Harfouche M, Aziz S, Ahmed A, Damseh R, et al. Wearable artificial intelligence for detecting anxiety: systematic review and meta-analysis. *J Med Internet Res*. Nov 08, 2023;25:e48754. [FREE Full text] [doi: [10.2196/48754](https://doi.org/10.2196/48754)] [Medline: [37938883](https://pubmed.ncbi.nlm.nih.gov/37938883/)]

## Abbreviations

**AI:** artificial intelligence

**EDA:** electrodermal activity

**HR:** heart rate

**HRV:** heart rate variability

**PRISMA-DTA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Diagnostic Test Accuracy

**QUADAS-2:** Quality Assessment of Diagnostic Accuracy Studies–Revised

**WESAD:** Wearable Stress and Affect Detection

*Edited by T de Azevedo Cardoso; submitted 10.09.23; peer-reviewed by M Zhou, Y Su, E Ikefuama; comments to author 19.10.23; revised version received 24.10.23; accepted 19.12.23; published 31.01.24*

*Please cite as:*

*Abd-alrazaq A, Alajlani M, Ahmad R, AlSaad R, Aziz S, Ahmed A, AlSahli M, Damseh R, Sheikh J*

*The Performance of Wearable AI in Detecting Stress Among Students: Systematic Review and Meta-Analysis*

*J Med Internet Res* 2024;26:e52622

URL: <https://www.jmir.org/2024/1/e52622>

doi: [10.2196/52622](https://doi.org/10.2196/52622)

PMID: [38294846](https://pubmed.ncbi.nlm.nih.gov/38294846/)

©Alaa Abd-alrazaq, Mohannad Alajlani, Reham Ahmad, Rawan AlSaad, Sarah Aziz, Arfan Ahmed, Mohammed AlSahli, Rafat Damseh, Javaid Sheikh. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org/>), 31.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.