

<https://doi.org/10.1038/s43856-025-01234-6>

# Wearable devices for anxiety assessment: a systematic review

Check for updates

Mohamed Elgendi<sup>1,2,8</sup>✉, Karmen Markov<sup>3,8</sup>, Hangyu Liu<sup>3,4</sup>, Maarten De Vos<sup>5</sup>, Kinda Khalaf<sup>1</sup>, Ahsan Khandoker<sup>1</sup>, Herbert F. Jelinek<sup>6</sup>, Debkalpa Goswami<sup>3,7</sup> & Carlo Menon<sup>3</sup>✉

## Abstract

**Background** There is growing interest in using biosignals from wearable devices to assess anxiety disorders. Among these, electrocardiography is the most widely used due to its ability to monitor cardiovascular activity. Other signals, such as respiratory, electrodermal activity, and photoplethysmography, also show promise. This review aims to evaluate how these signals, individually and in combination, have been used for anxiety detection.

**Methods** We systematically reviewed 26 studies published between 2014 and 2024 that used wearable devices to collect signals for anxiety detection. Extracted information included study design, signal types, features, classification methods, and accuracy outcomes. Pooled accuracies were calculated to compare single-signal and multi-signal approaches.

**Results** Here we show that approaches combining multiple signals outperform those using a single signal, with a pooled accuracy of 81.94% compared to 76.85%.

Electrocardiography was the most reliable individual signal, with a pooled accuracy of 80.34% across 12 studies. However, the limited number of single-sensor studies and methodological variability limit conclusions about the superiority of any one modality. The most common features included mean heart rate and heart rate variability for electrocardiography, the mean inspiratory-to-expiratory time ratio for respiratory signals, mean skin conductance for electrodermal activity, and the mean heart rate for photoplethysmography. Support vector machine was the predominant classifier.

**Conclusions** This review underscores the clinical potential of wearable devices for anxiety detection, emphasizing the value of multimodal approaches. Future research should focus on refining algorithms, expanding sample sizes, and exploring diverse contexts to improve the accuracy and generalizability of these methods.

## Plain Language Summary

Anxiety disorders are common, but detecting them in everyday life can be difficult.

Wearable devices, such as smartwatches, can record body signals that change with anxiety, including heart activity, breathing, skin responses, and blood flow. We reviewed 26 studies from the past decade that tested these signals for detecting anxiety. We found that combining several signals gives more accurate results than using a single signal, while heart activity was the most reliable single signal. However, most studies were small, used different devices and methods to trigger and measure anxiety, making it difficult to draw firm conclusions. Overall, wearable devices show strong potential for anxiety detection, but future research needs larger, standardized studies before they can be widely used in daily life and healthcare.

According to the World Health Organization, anxiety disorders (ADs) are the world's most common mental health challenge, affecting over 3.6% of the world's population, with worse impact in developing countries<sup>1</sup>. It is also estimated that, globally, one in seven (14%) of 10-19-year-olds experience mental health disorders, with depression, anxiety and behavioural disorders being the leading causes of illness and disability among adolescents<sup>2</sup>. The

continuously rising prevalence of AD is also of a great concern as demonstrated by the Global Burden of Disease report, which reflects a continuous increase in the last two decades between 2009 and 2019<sup>3</sup>. People suffering from ADs typically endure intense levels of fear or distress disproportionate to real events. ADs affect both adults and children. The anxiety/fear occupying the daily lives of these people affects their behaviors, thoughts, and

<sup>1</sup>Department of Biomedical Engineering and Biotechnology, Khalifa University of Science and Technology, Abu Dhabi, UAE. <sup>2</sup>Center for Biotechnology (BTC), Khalifa University of Science and Technology, Abu Dhabi, UAE. <sup>3</sup>Biomedical and Mobile Health Technology Lab, ETH Zurich, Zurich, Switzerland. <sup>4</sup>Department of Electrical Engineering and Information Technology, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. <sup>5</sup>ESAT - STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Department of Development and Regeneration, KU Leuven, Leuven, Belgium. <sup>6</sup>Department of Medical Sciences, Khalifa University of Science and Technology, Abu Dhabi, UAE. <sup>7</sup>Institute of Health Care Engineering with European Testing Center of Medical Devices, Graz University of Technology, Graz, Austria. <sup>8</sup>These authors contributed equally: Mohamed Elgendi, Karmen Markov. ✉e-mail: [mohamed.elgendi@ku.ac.ae](mailto:mohamed.elgendi@ku.ac.ae); [carlo.menon@hest.ethz.ch](mailto:carlo.menon@hest.ethz.ch)

physical health, often hindering their ability to function normally<sup>4</sup>. Several studies indicate a bidirectional relationship between anxiety disorders and cardiovascular diseases, i.e., anxiety disorders may be both causes and consequences of cardiovascular diseases<sup>5–8</sup>. The diagnosis of ADs is generally conducted in a clinic through a series of interviews and questionnaires. However, only one in four patients are reported to receive treatment for ADs due to lack of awareness, lack of competent or accessible mental health services, as well as lack of trained health care providers<sup>1</sup>. A conventional appointment with a practitioner at a clinic often results in delayed diagnosis, which can potentially lead to more severe illness<sup>9</sup>. Importantly, as mental health remains a stigma in many countries around the world, much less attention and resources are devoted to diagnostic and treatment tools for AD as compared to other medical conditions. While ADs are clinically diagnosed through structured assessments, experimentally induced anxiety in healthy individuals is often used to examine associated physiological responses. These controlled paradigms help uncover autonomic and physiological changes linked to anxiety. Understanding these responses is essential for developing objective, technology-driven screening methods.

Today's rapidly emerging wearable device (WD) technology provides an opportunity for a paradigm shift in precision medicine, including mental health applications, such as AD. This includes using technology-driven tools as effective, and affordable means for screening for ADs beyond traditional clinical settings. Commonly, individuals with ADs experience a wide range of physical symptoms spanning multiple physiological systems, including cardiovascular symptoms (i.e., accelerated heart rate (HR), palpitations, arrhythmia), and chest pain, respiratory symptoms (i.e., choking sensation, the sensation of a lump in the throat), and dyspnea, and neuro-muscular responses (i.e., stiffness, paresthesia, contractures, muscle tension, weakness, and fatigue)<sup>5,10</sup>. Current wearable technologies can quantify these changes and provide continuous, non-invasive monitoring capabilities in real time by leveraging various physiological signals, such as electrocardiography (ECG), respiratory signals (RSP), electrodermal activity (EDA), or photoplethysmography (PPG). More specifically, biosignal analysis provides the potential of detecting ADs by analyzing the different signals and identifying features and patterns depicting anomalies. Multimodal WDs integrating various measurement technologies can capture diverse physiological signals which significantly enhances reliability and accuracy. Additionally, implementing multimodal or multiplexed analysis techniques leads to fewer false positives and simultaneously provides multiple output signals that correlate to a particular physiological or pathophysiological state for active calibration and correction<sup>11</sup>. Importantly, WDs provide people with the ability to detect anxiety in its early stages and seek timely medical treatment, thus reducing the corresponding disease morbidity, mortality, and subsequent health-care costs.

ECG and PPG features, particularly heart rate variability (HRV), are widely used to detect anxiety disorders (ADs)<sup>12</sup>. ECG signals indicate the timing and rhythm of heartbeats by capturing the heart's electrical activity with high fidelity, enabling accurate derivation of HR and HRV metrics. However, ECG typically requires multi-electrode setups (e.g., chest or limb placement), which can limit wearability. In contrast, PPG captures peripheral blood volume changes using optical sensors, commonly worn at the wrist. Light is emitted into the skin, and the amount of light absorbed or reflected varies with blood flow during the cardiac cycle. PPG is widely used to derive cardiovascular parameters such as HR and oxygen saturation. While it offers better wearability, it is also more prone to motion artifacts and signal degradation under poor perfusion. Comparative studies show that PPG- and ECG-derived HRV metrics correlate well under ideal conditions, though PPG may be less reliable in ambulatory or real-world settings<sup>13,14</sup>. Despite decades of research on ECG-derived HRV, its utility in anxiety detection remains debated<sup>15</sup>. Respiration cycles involve inspiration (air inflow, diaphragm contraction, lung volume increase) and expiration (air outflow, diaphragm relaxation, lung volume decrease), producing measurable thoracoabdominal motion. Respiratory dysregulations, such as breath-to-breath respiratory instability, frequent sighing, are also common characteristics of ADs<sup>16</sup>. Breath rate generally increases under anxiety and

decreases under relaxation. Several studies indicated that ECG and RSP features were closely associated with anxiety and stress detection<sup>15,17</sup>. EDA reflects changes in skin conductance resulting from sweat gland activity, which is modulated by the autonomic nervous system. These changes occur in response to emotional arousal, stress, or cognitive load, making EDA a valuable marker for anxiety detection<sup>18,19</sup>. Some studies<sup>20–22</sup> also incorporate skin temperature (SKT) alongside EDA or PPG, as it may capture stress-related peripheral vasoconstriction. However, systematic assessments of these biosignals for anxiety detection are limited. Prior reviews<sup>23–26</sup> often focus more broadly on mental health rather than anxiety specifically, lack pooled performance comparisons across modalities or do not systematically assess study limitations.

This review examines 26 studies utilizing ECG, RSP, EDA, and PPG signals and their combinations, comparing single- and multimodal approaches. We analyze pooled accuracy, highlight the most frequently studied features and classifiers, and provide insights into the clinical potential of biosignal-based WDs for anxiety detection. We show that multimodal approaches achieve higher accuracy than single-signal methods, and that ECG is the most reliable individual signal. However, most studies are small and heterogeneous, limiting firm conclusions about the superiority of any one signal. Overall, wearable devices show strong potential for anxiety detection, but larger, standardized studies are needed to establish their role in clinical and everyday settings.

## Methods

### Search strategy and study eligibility

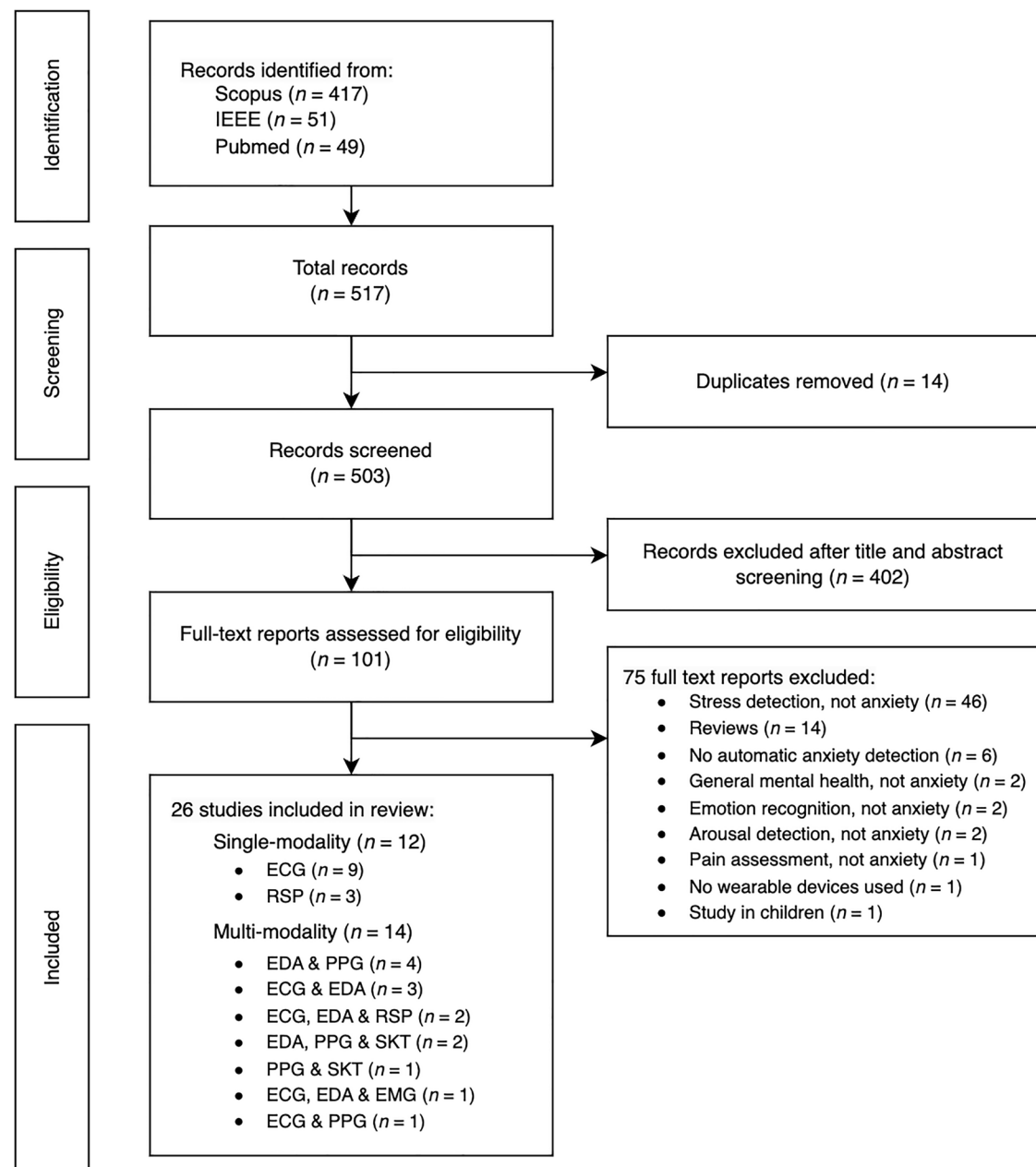
This systematic review was registered with the International Prospective Register of Systematic Reviews (PROSPERO, registration number CRD42025638476). Relevant peer-reviewed publications with full English text were selected in this review according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines<sup>27</sup>. Papers on anxiety detection using different signal inputs (ECG, RSP, EDA, PPG, and their combinations) were identified and reviewed. In the identification phase (see Fig. 1), three databases (Pubmed, IEEE, Scopus) were searched based on the following terms: (anxiety detection OR anxiety assessment) AND (electrocardiograph OR respiration signal OR respiratory signal OR breath OR EDA OR electrodermal activity OR PPG OR photoplethysmography) AND (wearable). While the search focused on ECG, RSP, EDA, and PPG, studies employing multimodal approaches that included additional signals (e.g. SKT or electromyography) alongside these core modalities were also included.

### Inclusion and exclusion criteria

This review selected studies published between 01 Jan 2014 and 01 Jan 2025, regardless of the study design and the country where the study was conducted. Before the screening, duplicate records were discarded from the identified papers based on the terms mentioned earlier. In the screening phase, the following inclusion criteria were applied: (1) Use of ECG, RSP, EDA, or PPG signals, or combinations, and their features used to detect anxiety; (2) Clear quantification/description of performance of methods and signals; (3) Use of WDs to obtain noninvasive biosignals; (4) Studies conducted in adults. Review articles and publications not written in English were excluded.

### Data extraction

The data extraction process focused on collecting comprehensive information from each publication to enable meaningful comparisons, including signal modalities utilized, anxiety induction methods, anxiety measurement tools, WDs used and their placement, sample sizes of the studies, extracted features, best-performing machine learning methods, validation methods for machine learning models, and evaluation metrics. While all reported metrics (e.g., precision, recall, F1-score) were extracted, accuracy (ACC) was used as the primary metric to ensure consistency, due to being the most consistently reported. However, it is important to note that ACC can be biased when class distributions are imbalanced, as it does not account for



**Fig. 1 | PRISMA diagram showing study selection process.** The diagram illustrates the study identification, screening, eligibility assessment, and inclusion process. A total of 517 records were retrieved from Scopus, IEEE, and PubMed, of which 14 duplicates were removed. After title and abstract screening, 402 records were excluded. Of 101 full-text reports assessed, 75 were excluded for reasons including

stress detection (not anxiety), reviews, absence of automatic anxiety detection, non-anxiety outcomes, non-wearable devices, or studies in children. Finally, 26 studies were included in the review. ECG electrocardiogram, EDA electrodermal activity, EMG electromyography, PPG photoplethysmography, RSP respiratory signal, SKT skin temperature.

false positives or false negatives in the minority class. Metrics like F1-score or Matthews correlation coefficient (MCC) better address this by incorporating both precision and recall, providing a more balanced view of model performance under class imbalance.

ACC is calculated as follows (see Eq. (1)):

$$ACC = \frac{TP + TN}{TN + TP + FN + FP}, \quad (1)$$

where:

- TP (True Positives) denotes the number of anxiety cases correctly identified and classified,

- TN (True Negatives) denotes the number of non-anxiety cases correctly classified,
- FN (False Negatives) represents the number of anxiety cases incorrectly classified as non-anxiety, and
- FP (False Positives) indicates the number of non-anxiety cases incorrectly classified as anxiety.

For studies that did not explicitly report accuracy, we estimated it using the provided TP, TN, FN, and FP values from that study. This estimated accuracy, calculated using the equation mentioned above, is denoted by ACC. This is distinct from ACC, which indicates that the accuracy was directly reported in the study.

## Pooled accuracy calculation

To determine the pooled accuracy across various datasets or studies, we employed the Cochrane formula<sup>28,29</sup> for weighted averages. In this approach, the accuracy of each study is weighted by its sample size, and the sum is then divided by the aggregate sample size from all the studies. The equation for pooled accuracy is as follows:

$$\text{Pooled ACC} = \frac{\sum_{i=1}^n (N_i \times \text{ACC}_i)}{\sum_{i=1}^n N_i}, \quad (2)$$

where:

- $N_i$  is the sample size of study  $i$ ,
- $\text{ACC}_i$  is the accuracy of study  $i$ , and
- $n$  is the total number of studies.

## Statistics and reproducibility

Statistical analyses focused on ACC as the primary metric. Details of ACC and pooled ACC calculations, including weighting by sample size, are provided in the sections Data extraction and Pooled accuracy calculation. Reproducibility was ensured by applying predefined inclusion and exclusion criteria, and by extracting study characteristics systematically using a standardized template. Sample sizes were taken directly from each included study, and each study was considered one independent unit of analysis (replicate) in the synthesis. A completed PRISMA checklist is provided in the Supplementary Material to document compliance with reporting standards.

## Limitation assessment criteria

The limitations of each included study were systematically assessed across eight predefined categories: population representation (L1), sample size (L2), study completion rate (L3), selective reporting (L4), outcome measurement validity (L5), algorithmic transparency (L6), validation robustness (L7), and reproducibility (L8). Each category was assigned a level (low, moderate, serious, or could not be determined) based on predefined criteria. For example, population representation (L1) was evaluated based on demographic diversity, with homogeneous samples (e.g., single gender or narrow age groups) rated as serious and diverse populations rated as low. Sample size (L2) thresholds were set at <30 participants (Serious), 30–99 participants (moderate), and ≥100 participants (low). Study completion rate (L3) considered dropout documentation, while selective reporting (L4) evaluated comprehensiveness in presenting evaluation metrics and subgroup analyses. Outcome measurement validity (L5) focused on whether validated tools or unvalidated self-reports were used for anxiety assessment. Algorithmic transparency (L6) evaluated whether key methodological details, such as extracted features, classifier settings, and preprocessing steps, were sufficiently described. Validation robustness (L7) examined the use of robust techniques such as leave-one-subject-out (LOSO) cross-validation (CV) or external datasets, with weaker methods (e.g., training/testing overlap) rated as serious. Finally, reproducibility (L8) was assessed based on the availability of datasets, code, and tools for independent replication. This framework, developed for the present study, provided a consistent framework to evaluate limitations across studies.

## Results

### Study selection

As shown in Fig. 1, we identified 517 records from three databases: Scopus ( $n = 417$ ), IEEE ( $n = 51$ ), and PubMed ( $n = 49$ ). Upon removing 14 duplicates, 503 studies remained for screening. Following title and abstract screening, 402 studies were excluded. The remaining 101 full-text reports were assessed for eligibility, leading to the exclusion of additional 75 studies based on the following reasons: stress detection, not anxiety ( $n = 46$ ), review articles ( $n = 14$ ), no automatic anxiety detection ( $n = 6$ ), general mental health status, not anxiety ( $n = 2$ ), emotion recognition, not anxiety ( $n = 2$ ), arousal detection, not anxiety ( $n = 2$ ), pain assessment, not anxiety ( $n = 1$ ), no use of WDs ( $n = 1$ ), and studies conducted in children ( $n = 1$ ).

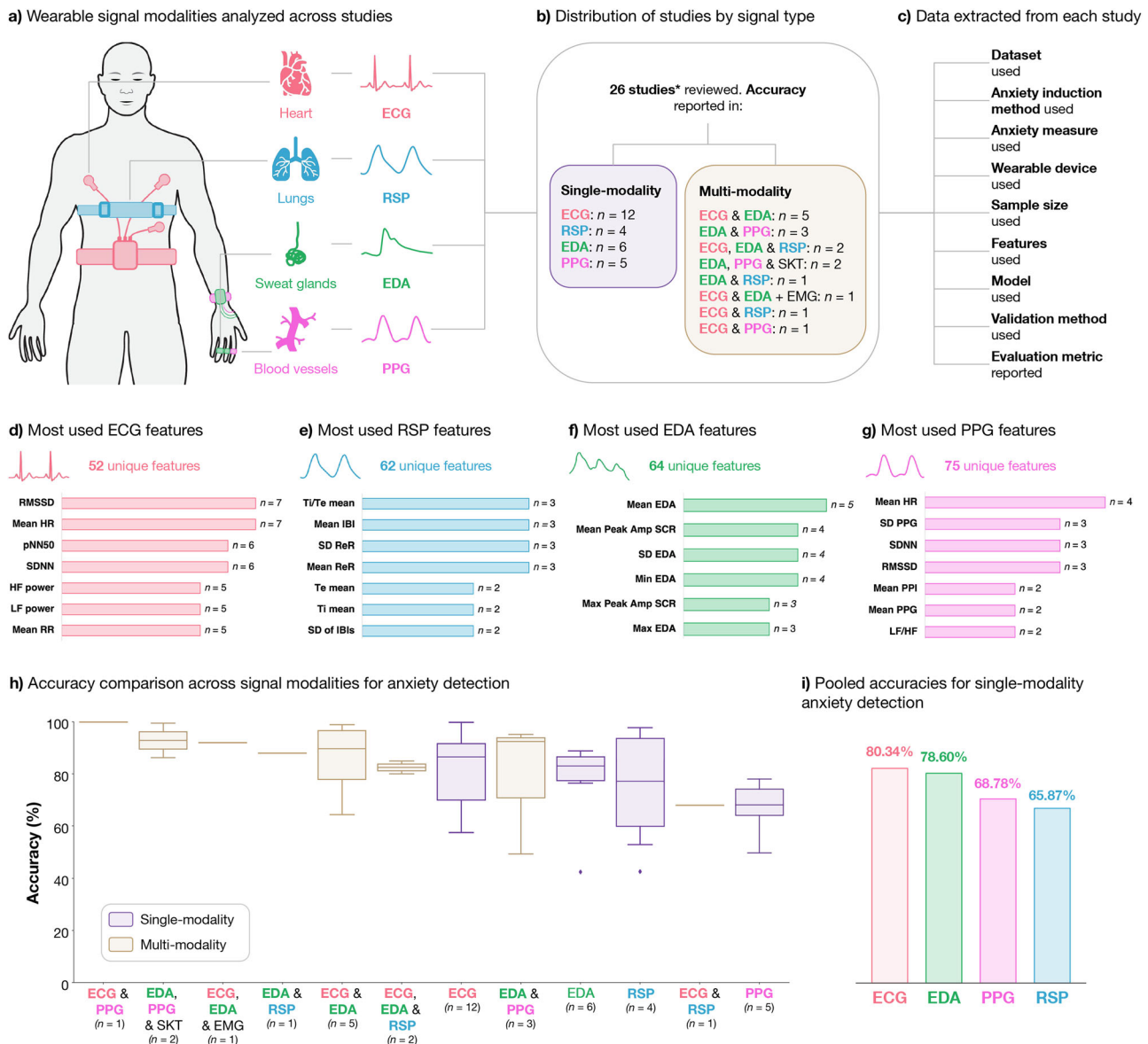
Ultimately, 26 studies were included in the review. Of these, 12 studies used single-modality signals, including ECG ( $n = 9$ )<sup>30–38</sup>, RSP ( $n = 3$ )<sup>39–41</sup>. The remaining 14 studies employed multi-modality approaches or reported results for multiple signals: EDA & PPG ( $n = 4$ )<sup>18,19,42,43</sup>, ECG & EDA ( $n = 3$ )<sup>44–46</sup>, ECG, EDA & RSP ( $n = 2$ )<sup>47,48</sup>, EDA, PPG & SKT ( $n = 2$ )<sup>20,21</sup>, PPG & SKT ( $n = 1$ )<sup>22</sup>, ECG, EDA & EMG ( $n = 1$ )<sup>49</sup>, and ECG & PPG ( $n = 1$ )<sup>50</sup>. The study design and data analysis workflow for the 26 included studies are summarized in Fig. 2. By providing an overview of the key methodological steps, this figure highlights the diversity in study designs and approach among the 26 articles, which is further elaborated in the subsequent sections.

### Included studies

The 26 studies reviewed (summarized in Supplementary Data 1 and Supplementary Data 2) employed a variety of biosensors, datasets, anxiety induction and measurement methods, data preprocessing and machine learning approaches, resulting in a wide range of reported accuracies. The majority of these studies were conducted in controlled lab environments, with only four studies<sup>37–39,41</sup> utilizing real-world settings. For single biosensors, classification accuracies ranged from 42.53% for EDA, reported by Lee et al.<sup>18</sup>, to 99.95% for ECG by Baygin et al.<sup>30</sup>. For multisensor recordings, 99.84% was the highest reported accuracy by Sinche et al.<sup>50</sup>, based on ECG and PPG sensor recordings, while Lee et al.<sup>18</sup> reported the lowest accuracy of 49.26% when combining PPG, EDA, and EEG recordings. The differences in accuracies have to be viewed with caution as they are derived from different recording devices (e.g., Empatica E4<sup>18</sup>, SS2LB ECG module<sup>30</sup>, AD8232 and MAX30100 sensors<sup>50</sup>), sample sizes, machine learning algorithms, and study conditions, which should be considered when comparing results. Incorporating contextual features, such as demographic and situational data, has been shown to enhance classification accuracies in anxiety detection. For instance, Jain & Kumar<sup>44</sup> utilized the Wearable Stress and Affect Detection (WESAD) dataset<sup>51</sup> to classify anxiety levels in 15 participants using ECG and EDA signals from a chest-worn RespiBAN device. They achieved accuracies of 89.80% for ECG, 85.90% for EDA, and 96.70% for combined signals, which further increased to 97.30% with the addition of contextual features. Similarly, Nath & Thapliyal<sup>49</sup> analyzed data from 41 older adults using wristband sensors to capture EDA and PPG signals during the Trier Social Stress Test (TSST)<sup>52</sup>, reporting accuracies of 89.00% for EDA and 78.00% for PPG. The inclusion of contextual features improved these accuracies to 92.00% for EDA and 83.00% for PPG. However, several studies reported accuracies exceeding 90% also without contextual features. For example, Baygin et al.<sup>30</sup>, Sinche et al.<sup>50</sup>, Shaikat-Jali et al.<sup>21</sup>, Zhou et al.<sup>45</sup>, Banerjee et al.<sup>40</sup>, Jain & Kumar<sup>44</sup>, Padmaja et al.<sup>32</sup>, Di Tecco et al.<sup>42</sup>, Petrescu et al.<sup>43</sup>, Haritha et al.<sup>41</sup>, Tripathy et al.<sup>31</sup>, Vaz et al.<sup>49</sup>, Wen et al.<sup>35</sup> achieved notable results. Among these, Vaz et al.<sup>49</sup> achieved an accuracy of 84.50% using ECG, EDA, and EMG signals from the WESAD dataset<sup>51</sup>. To address the issue of class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) was applied. This led to a notable increase in accuracy to 92.00%. SMOTE's ability to enhance performance highlights the importance of addressing dataset imbalances in developing robust anxiety detection models.

### Datasets employed across studies

The reviewed studies employed a variety of datasets for anxiety detection. A significant proportion, 16 studies<sup>18–22,32,35–43,50</sup>, collected data specifically for their research. The WESAD dataset<sup>51</sup>, used in three studies<sup>44,45,49</sup>, comprises of data from 15 participants collected using ECG, EDA, and respiratory signals captured using a chest-worn RespiBAN device. Anxiety was induced through the TSST<sup>52</sup>, and anxiety levels were categorized as low, moderate, or high based on self-reports from the 6-item State and Trait Anxiety Inventory (6-STAI)<sup>53</sup>. The dataset by Ihmig et al. 2020<sup>54</sup>, used in four of the reviewed studies<sup>33,46–48</sup>, includes ECG, respiratory (RSP), and skin conductance signals from 57 participants aged 18 to 40 years. The signals were recorded using a BITalino device during a protocol involving 16 spider-themed video clips presented in a virtual reality (VR) setting. Each session consisted of four 1 min spider-themed clips, followed by a 5-minute resting period.



**Fig. 2 | Study design and data analysis workflow for wearable-based anxiety detection review.** The study design and data analysis workflow consisted of four steps: search and selection of studies, data extraction, feature analysis, and performance analysis. **Step 1: Search and selection of studies.** **a** shows the wearable signal modalities analyzed: ECG, RSP, EDA, and PPG. **b** presents the number of reviewed studies using each signal modality, including single- and multi-modality approaches. Some studies contributed to multiple categories; for example, Jain & Kumar<sup>44</sup> reported results using ECG alone, EDA alone, and ECG + EDA combined, and were counted in all relevant categories. **Step 2: Data extraction.** **c** highlights the elements extracted from each study, including dataset, features, model, validation, and evaluation metrics. **Step 3: Feature analysis.** **d–g** summarize the most commonly used features for ECG, RSP, EDA, and PPG, respectively, and their frequencies of use. **Step 4: Performance analysis.** **h** displays boxplots illustrating the distribution of accuracy results across different signal modalities, comparing single-modality and

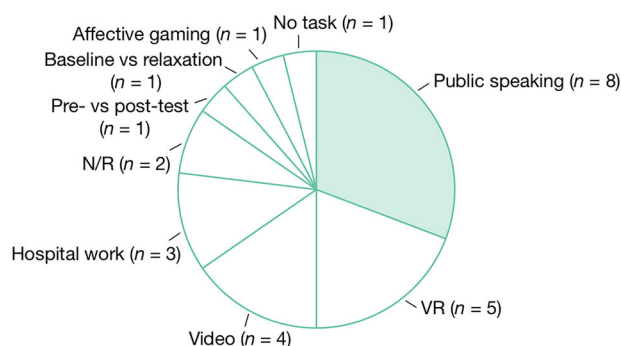
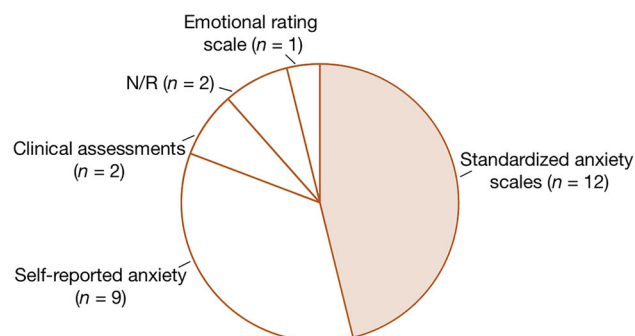
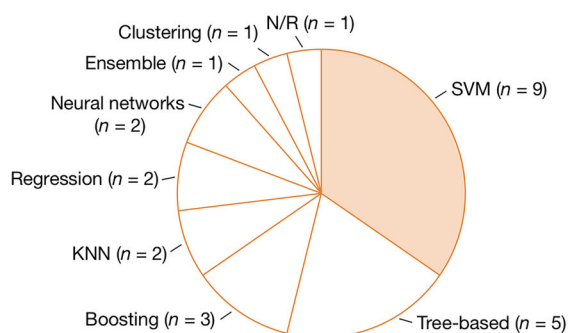
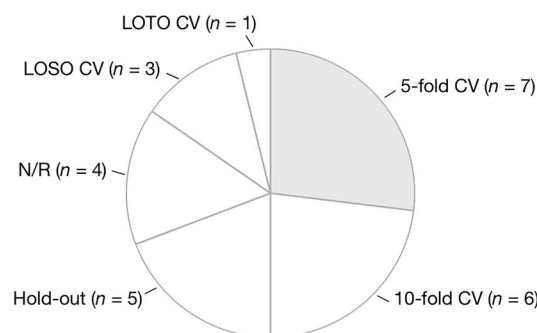
multi-modality approaches, while **i** presents pooled accuracies for single-modality anxiety detection, highlighting the overall performance of ECG, EDA, PPG, and RSP signals. **Note:** In **(b)**,  $n$  refers to the number of independent studies using each signal modality (single- or multi-modality). In **d–g**,  $n$  refers to the number of independent studies that reported the use of each specific feature. In **(h)**,  $n$  indicates the number of independent study results contributing to each boxplot. ECG electrocardiogram, EDA electrodermal activity, PPG photoplethysmography, RSP respiratory signal, HR heart rate, SDNN standard deviation of normal-to-normal intervals, RMSSD root mean square of successive differences, HF power high-frequency power, LF power low-frequency power, Mean RR mean R-R interval, Ti/Te inspiratory to expiratory time ratio, IBI inter-breath interval, ReR respiratory rate, SCL skin conductance level, SCR skin conductance response, PPI pulse interval, LF/HF low-frequency to high-frequency power ratio.

Participants rated their subjective arousal on a four-point scale ("1 = not at all" to "4 = strongly") after each clip. The dataset by Elgendi et al. 2022<sup>55</sup>, used in two studies<sup>30,31</sup>, includes ECG signals collected from 19 participants exposed to anxiety-inducing and non-anxiety video clips. Anxiety levels were measured using the Hamilton Anxiety Rating Scale (HAM)<sup>56</sup>, and the signals were recorded using the SS2LB ECG module. The Anxiety Phases Dataset (APD)<sup>57</sup>, used in Zhou et al. 2023<sup>45</sup>, includes data from 52 participants who performed public speaking and bug-box tasks. Anxiety levels

were assessed using the Subjective Units of Distress Scale (SUDS)<sup>58</sup>. Signals were recorded using the Zephyr BioHarness 3.0 for chest ECG and the Grove-GSR Sensor for wrist EDA. Lastly, an undefined Kaggle dataset focusing on ECG signals was used in Tang et al. 2021<sup>34</sup>.

#### Anxiety induction methods employed across studies

The studies reviewed here employed various anxiety induction methods, as illustrated in Fig. 3a. Public speaking tasks, which were the most common

**a) Anxiety induction methods****b) Anxiety measurement methods****c) Models used for anxiety detection****d) Validation methods****Fig. 3 | Overview of anxiety detection methodologies across reviewed studies.**

**a** Anxiety induction methods: Proportions of methods used to induce anxiety in the studies. **b** Anxiety measurement methods: Proportions of methods used to measure anxiety across studies. **c** Models used for anxiety detection: Proportions of machine learning models employed for anxiety detection. **d** Validation methods: Proportions

of strategies employed to validate models. **Note:** In all panels, *n* refers to the number of independent studies reporting the use of the respective method. LOTO CV leave-one-trial-out cross-validation, LOSO CV leave-one-subject-out cross-validation, N/R not reported, VR virtual reality, SVM support vector machines, KNN K-Nearest Neighbors, CV cross-validation.

method, were used in eight studies. These included the TSST<sup>52</sup>, which involves public speaking and mental arithmetic task, used in five studies<sup>19,35,44,45,49</sup>; a public speaking task with a bug-box component<sup>45</sup>, an impromptu speech task<sup>21</sup>; and a Virtual Reality Exposure Therapy (VRET) public speaking scenario<sup>20</sup>. VR-based anxiety induction methods were employed in five studies, including spider exposure<sup>33,46–48</sup> and exposure to heights<sup>13</sup>. Video-based methods were used in four studies, including general anxiety-inducing videos<sup>30,31</sup>, a purposely edited horror movie trailer<sup>12</sup>, and driving anxiety-inducing videos<sup>18</sup>. Other anxiety induction techniques included detecting anxiety in hospital shift work settings<sup>37–39</sup>, pre- and post-test evaluations comparing anxious versus non-anxious states<sup>50</sup>, baseline versus relaxation tasks with biofeedback for prenatal anxiety detection in pregnant women<sup>22</sup>, an affective Pacman gaming task<sup>59</sup> to induce frustration<sup>40</sup>, and an experimental task designed specifically for Social Anxiety Disorder (SAD)<sup>36</sup>. Additionally, two studies did not specify the anxiety induction methods used<sup>32,34</sup>.

**Anxiety measurement tools employed across studies**

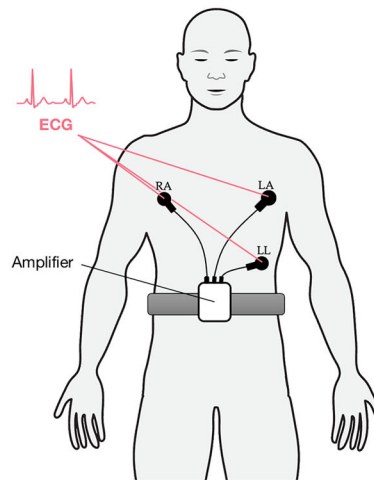
The reviewed studies utilized a range of tools to measure the levels of anxiety, encompassing both standardized scales and self-reported measures, as illustrated in Fig. 3b. Standardized anxiety scales were the most commonly used method, applied in 12 studies, including STAI<sup>53</sup> used in five studies<sup>19,22,44,45,49</sup>, SUDS<sup>58</sup>, which was used in three<sup>20,43,45</sup>, Cognitive Test Anxiety Scale<sup>60</sup> used in one<sup>50</sup>, Social Phobia Screening Questionnaire (SPSQ)<sup>61</sup> used in one<sup>21</sup>, HAM<sup>56</sup> used in two<sup>30,31</sup>, and LSAS<sup>62</sup> used in two studies<sup>21,36</sup>. Self-reported anxiety levels were employed in nine studies, with methods ranging from self-reported anxiety onset<sup>18,48</sup>, predominant emotion<sup>42</sup>, and daily ratings on a 5-point scale<sup>37–39</sup> to general self-reported

anxiety levels<sup>33,35,46,47</sup>. Additionally, Wen et al.<sup>35</sup> employed audience-observed anxiety scores based on symptoms, such as sweating, trembling, blushing and avoiding eye contact as ground truths for labeling anxiety states. Clinical assessments were used in two studies to measure anxiety, including the Diagnostic and Statistical Manual of Mental Disorders (DSM-5)<sup>63</sup> diagnostic criteria in<sup>36</sup>, and clinical psychiatric diagnoses in<sup>41</sup>. Emotional rating scales, such as Self-Assessment Manikin (SAM), were employed in a single study<sup>40</sup>. Two studies<sup>32,34</sup> did not report specific anxiety measurement tools, which limits the comparability of their findings. Overall, the studies highlight the diverse methods used to measure anxiety.

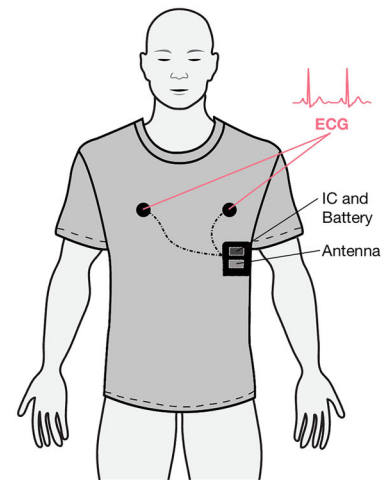
In terms of classification schemes, anxiety was labeled as binary (e.g., anxious vs. non-anxious) in 18 studies<sup>18,19,21,32–35,37–41,45–50</sup>, while four studies used 3-class schemes (e.g., low/moderate/high)<sup>22,42–44</sup>, and three studies used 4-class schemes (e.g., normal/mild/moderate/severe)<sup>20,30,31</sup>. One study did not report its labeling method<sup>36</sup>. Most studies that used standardized scales also used them to define class labels. However, in some cases, labels were derived from experimental context. For example, Vulpe-Grigorasi et al.<sup>33</sup>, Gazi et al.<sup>47</sup>, and Khullar et al.<sup>48</sup> assigned labels based on exposure to spider-related VR clips versus rest, despite collecting self-reported anxiety levels. Sinche et al.<sup>50</sup> used the Cognitive Test Anxiety Scale for assessment, but labeled data based on timing, pre-test as anxious and post-test as non-anxious. Banerjee et al.<sup>40</sup> used SAM ratings, but labeled gameplay windows as anxious or non-anxious depending on whether participants played a frustration-inducing or standard Pacman game. Finally, Shaikat-Jali et al.<sup>21</sup> employed LSAS-SR and SPSQ for anxiety assessment, but classification into anxious versus baseline was based on the timing of data collection relative to the impromptu speech task.

**Fig. 4 | Placement of ECG and RSP WDs in this study.** **a** Wearable ECG device. RA is right arm lead, LA is left arm lead and LL is left leg lead. **b** T-shirt ECG device. **c** Inductive RSP device. **d** Piezoelectric RSP device. **e** Wrist-mounted EDA and PPG device. **f** Finger-mounted EDA and PPG device. ECG electrocardiogram, EDA electrodermal activity, PPG photoplethysmography, RSP respiratory signal.

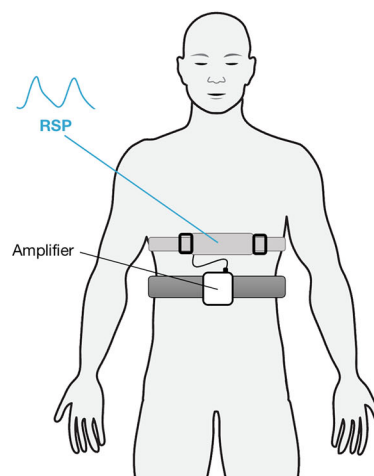
**a) Wearable ECG device**



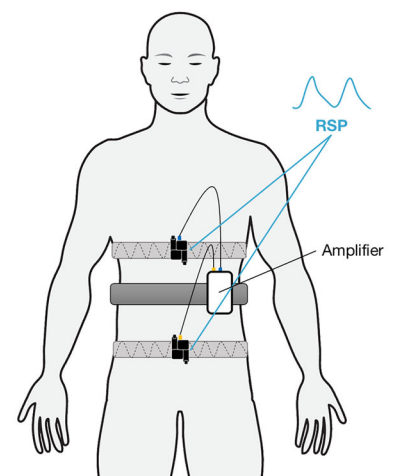
**b) T-shirt ECG device**



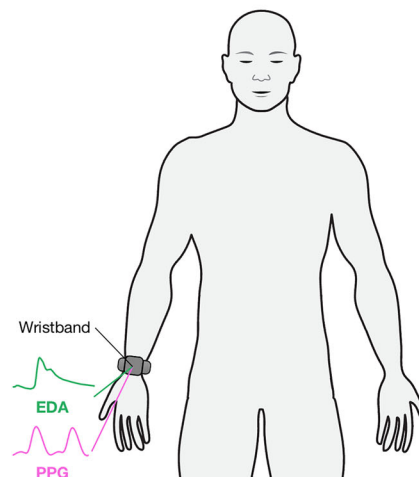
**c) Inductive RSP device**



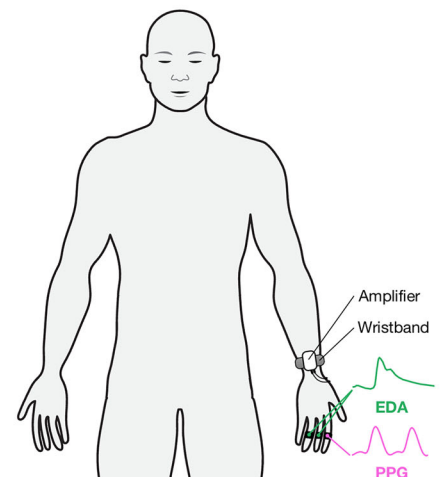
**d) Piezoelectric RSP device**



**e) Wrist-mounted EDA and PPG device**



**f) Finger-mounted EDA and PPG device**



### Wearable devices employed across studies

The reviewed studies included here used various WDs with diverse configurations and placements to capture ECG, EDA, PPG, and RSP signals for anxiety detection. The sketches of the devices used are provided in Fig. 4, and the devices that each study used are listed in Supplementary Data 1 and Supplementary Data 2. Below, we group the

devices based on the signal modalities they recorded in the included studies.

Nine devices were employed to record single signal modalities in the studies reviewed. Out of these, five devices were used to record ECG signals: (1) the SS2LB ECG module<sup>30,31</sup>; (2) AD8232 sensor<sup>50</sup>; (3) the Biopac MP-45 system<sup>32</sup>; (4) the MP150 recorder<sup>35</sup>; (5) the Zephyr BioHarness 3.0<sup>45</sup>. The

SS2LB ECG module, used in the Elgendi et al.<sup>55</sup> dataset utilized by Baygin et al.<sup>30</sup> and Tripathy et al.<sup>31</sup> employed a Lead I configuration (right arm (RA) to left arm (LA), left leg (LL) as ground), designed for high-resolution recordings compatible with Biopac systems. The AD8232 sensor, used by Sinche et al.<sup>50</sup>, consists of a compact analog front-end module placed on the left side of the chest with three electrodes configured in a standard Lead II setup. The Biopac MP-45 system, used by Padmaja et al.<sup>32</sup>, featured an embedded CPU for data acquisition and supported various electrode types, though specific placements were not detailed. The MP150 multi-channel recorder, used by Wen et al.<sup>35</sup>, collected ECG signals with electrodes placed on both wrists and the right ankle, sampling at 400 Hz, alongside skin conductance data. Lastly, the Zephyr BioHarness 3.0, used in the APD<sup>57</sup> dataset utilized by Zhou et al.<sup>45</sup>, recorded chest-based ECG data with two disposable adhesive electrodes aligned to the breastbone, sampling at 250 Hz. Most of these devices, as illustrated in Fig. 4a, are attached to the subject with a strap. Two devices were employed to record RSP data: (1) Embletta<sup>41</sup> and (2) an undefined respiration belt<sup>40</sup>. The Embletta device (Fig. 4c), used by Haritha et al.<sup>41</sup>, contains two elasticity belts with sinusoid conducting coils for thoracic-abdominal movement detection. One belt is placed on the chest, and the other is placed on the subject's abdomen, such that variations of the thoracic and abdominal volumes cause changes in the surface of the section and stretching of elasticity belts. This leads to variation of the self-inductance of the coils and the frequency of their oscillation, hence inducing a current<sup>64</sup>. Banerjee et al.<sup>40</sup> obtained the RSP using a respiration belt that measures the expansion of the abdomen<sup>65</sup> related to the quantity of inspired and expired air. One sensor was used to measure EDA signal, the Grove-GSR Sensor V1.2, utilized by Zhou et al.<sup>45</sup>. It measured EDA signals at a sampling rate of 50 Hz using electrodes placed on the index and ring fingers, with the device enclosed in a 3D-printed module secured on both wrists via Velcro straps. One sensor, the MAX30100 sensor, employed by Sinche et al.<sup>50</sup>, recorded PPG signals from the students' index finger.

Five devices were employed to record multiple signal modalities in the studies reviewed: (1) the RespiBAN device<sup>44,45,49</sup> for recording ECG and EDA signals; (2) the OMSignal smartshirt<sup>37–39</sup> for ECG and RSP signals; (3) BiTalino<sup>33,46–48</sup> for ECG, EDA and RSP signals; (4) Empatica E4<sup>18,20–22,44,45,49</sup> for EDA and PPG signals; (5) Shimmer3 GSR+<sup>42,43</sup> also for EDA and PPG signals. The RespiBAN device, a chest-worn sensor capable of capturing ECG, EDA, EMG, RSP, body temperature, and acceleration, was used in the WESAD dataset<sup>51</sup> and employed by Jain and Kumar<sup>44</sup>, Zhou et al.<sup>45</sup>, and Vaz et al.<sup>49</sup>. The OMSignal smartshirt (Fig. 4b) used by Tiwari et al.<sup>37–39</sup> contains a close-fitting T-shirt and an OMBox. The T-shirt has sensors that are directly knit into the textile. It leverages conductive yarns with electromechanical properties. The OMBox is fastened to the T-shirt using snap buttons. A battery that lasts for 30 h of real-time usage powers the OMBox. The built-in Bluetooth module connects to the mobile phone and transfers data. As shown in Fig. 4d, the BiTalino respiration belt, which was used in the dataset by Ihmig et al. 2020<sup>54</sup> and used in four studies<sup>33,46–48</sup> is strapped around the subject's chest for recording changes in thoracic movement during breathing. The Empatica E4 wristband (Fig. 4e), used in the WESAD dataset<sup>51</sup> utilized in three studies<sup>44,45,49</sup> and four further studies<sup>18,20–22</sup>, is worn on the non-dominant wrist and records EDA at 4 Hz, PPG at 64 Hz, body temperature at 4 Hz, and three-axis acceleration at 32 Hz. The Shimmer3 GSR+ device (Fig. 4f), used in two reviewed studies<sup>42,43</sup>, measures EDA using electrodes placed on the index and middle fingers, capturing PPG signals via an optical pulse probe on the ring finger and transmitting data in real-time via Bluetooth.

## Features employed across studies

For ECG signals, a total of 52 unique features were extracted across the reviewed studies (Fig. 5a). The mean HR<sup>36,43–47,50</sup> and RMSSD<sup>33,37,38,44,45,47,50</sup> were the most frequently utilized features, both used in seven studies. The other frequently used features included standard deviation of RR intervals (SDNN), also used in six studies<sup>35,37,44,45,47,50</sup>; as well as the percentage of successive RR intervals differing by more than 50 ms (pNN50), used in six

studies<sup>35,37,38,47,49,50</sup>. Additionally, low- (LF) and high-frequency (HF) power of RR intervals were extracted in five studies<sup>34,35,37,38,45</sup>, and the mean RR interval was also employed in five studies<sup>35,37,38,44,47</sup> (Fig. 5b).

For PPG signals, 75 unique features were extracted across the reviewed studies (Fig. 5a). The most frequently used feature was the mean HR, which appeared in four different studies<sup>19,20,22,42</sup>. The standard deviation of the PPG signal (SD PPG)<sup>20,22,42</sup> and the standard deviation of the NN intervals (SDNN)<sup>20,22,50</sup>, as well as the root mean square of successive differences of NN intervals (RMSSD)<sup>20,22,50</sup> were reported in three studies. Other extracted features that were reported in more than one study included the mean inter-beat interval (Mean PPI)<sup>20,50</sup>, the mean of the PPG signal (Mean PPG)<sup>20,22</sup>, and the low-frequency to high-frequency power ratio (LF/HF)<sup>18,22</sup> (Fig. 5c).

For EDA signals, 62 unique features were extracted across the reviewed studies (Fig. 5a). The most frequently employed feature was the mean of the EDA signal (Mean EDA), utilized in five studies<sup>18,20,44–46</sup>. Other commonly reported features included the mean peak amplitude of the phasic skin conductance response (SCR), the phasic component of EDA (Mean Peak Amp SCR), used in four studies<sup>19,20,44,47</sup>, and the standard deviation of the EDA signal (SD EDA), also reported in four studies<sup>18,20,44,45</sup>. The minimum (Min EDA) and maximum (Max EDA) values of the EDA signal were each utilized in three studies<sup>18,20,44</sup>. Additionally, the maximum amplitude of the detected peaks in the SCR (Max Peak Amp SCR) was also extracted in three studies<sup>19,20,49</sup> (Fig. 5d).

For RSP signals, 62 unique features were extracted across the reviewed studies (Fig. 5a). The most commonly utilized features included the mean ratio of inspiratory time to expiratory time (Ti/Te mean)<sup>39,40,47</sup>, and also reported in the same three studies<sup>39,41,47</sup>, the mean breath-by-breath inter-breath intervals (Mean IBI), the mean respiratory rate (Mean ReR), and the standard deviation of respiratory rate (SD ReR). Other notable features found included the mean inspiratory time (Ti mean) and mean expiratory time (Te mean), both reported in two studies<sup>40,47</sup>, as well as the standard deviation of breath-by-breath interbreath intervals (SD of IBIs), reported in two studies<sup>39,47</sup> (Fig. 5e).

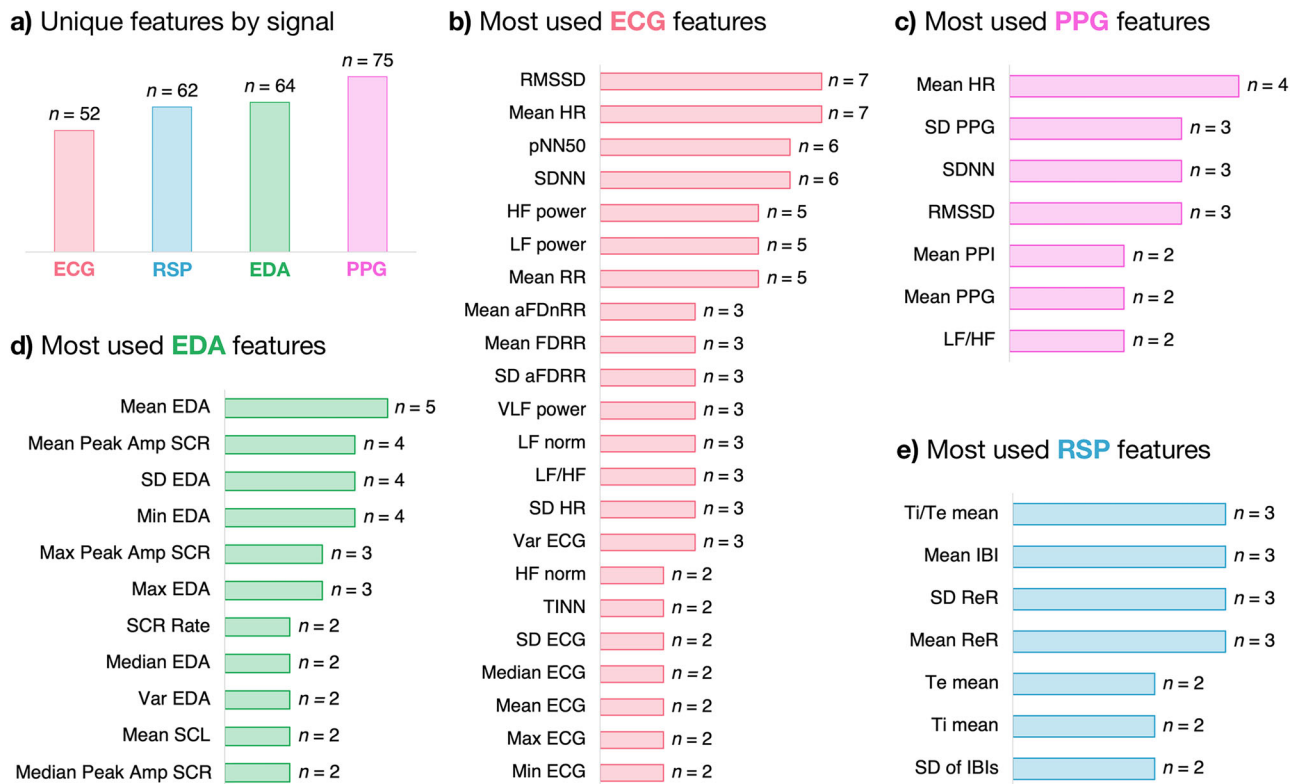
## Models and validation methods

The studies reviewed here employed a variety of machine learning models to detect anxiety as illustrated in Fig. 3c. Among these, Support Vector Machines (SVMs) were the most commonly used models, employed in nine studies<sup>20,22,30,35,37–39,41,50</sup>. Tree-based models were also widely used, with five studies using methods such as Random Forest (RF)<sup>19,47</sup>, Decision Tree (DT)<sup>36</sup>, ET<sup>48</sup>, and Bagged Trees<sup>46</sup>. Boosting models were used in three studies, including Adaptive Boosting<sup>49</sup>, Extreme Gradient Boosting (XGB)<sup>31</sup>, and Gradient Tree Boosting (GTB)<sup>44</sup>. K-Nearest Neighbors (KNN) models were applied in two studies<sup>21,42</sup> and regression models were also used in two<sup>18,43</sup>. Two other studies explored neural network architectures, including using a Multi-Layer Perceptron (MLP) model<sup>40</sup> and a 1D Convolutional Neural Network (CNN)<sup>33</sup>. Other methods included ensemble techniques, combining SVM, Light Gradient Boosting Machine (LGBM), RF, and XGB<sup>45</sup>, and clustering-based methods, such as a Density-Based Spatial Clustering of Applications with Noise (DBSCAN)-enhanced algorithm<sup>34</sup>. Finally, one study<sup>32</sup> did not report the model used.

The most common validation approach was 5-fold CV, used in seven studies<sup>35,37–39,45,48,49</sup>, followed by 10-fold CV in six studies<sup>20,21,30,40,46,50</sup>. Hold-out validation was used in five studies<sup>19,31,33,42,43</sup>. LOSO CV was applied in three studies<sup>22,44,47</sup>, while Leave-One-Task-Out (LOTO) CV was employed in one study<sup>18</sup>. Four studies<sup>32,34,36,41</sup> did not specify the validation methods used (Fig. 3d).

## Pooled and mean accuracies

The studies reviewed in this article reported varying levels of accuracy across single-modality and multi-modality approaches for capturing and analyzing various physiological signals reflecting AD, as summarized in Fig. 6a, b, and Table 1. It is important to note that mean and pooled accuracies were calculated only from studies which reported results on accuracy, or where accuracy was possible to calculate. For example, the results from Bao et al.<sup>22</sup>



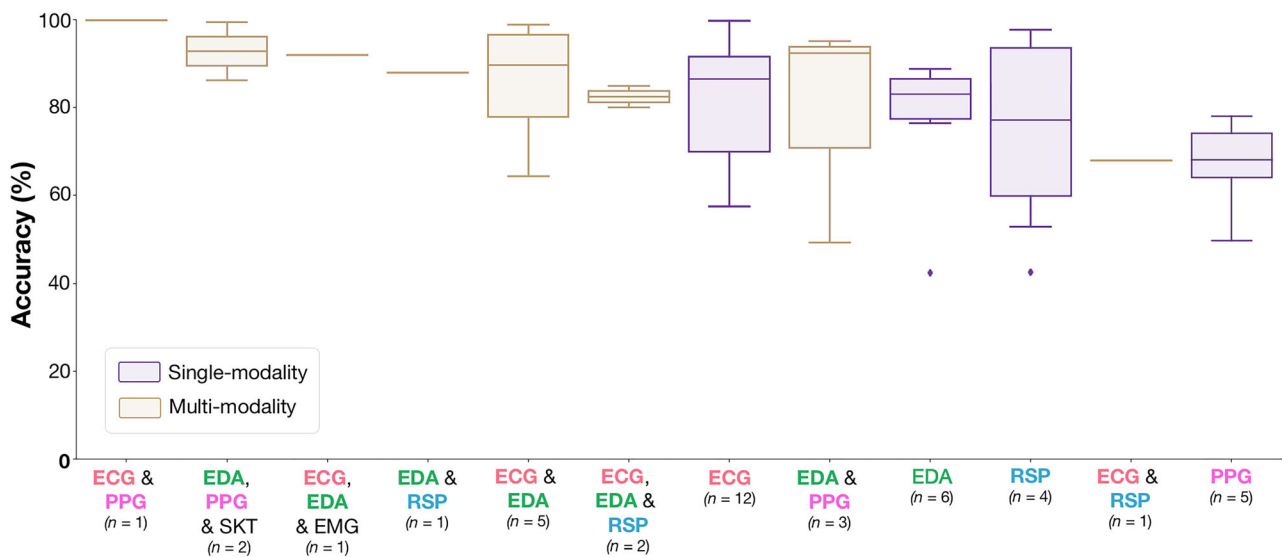
**Fig. 5 | Overview of features extracted from physiological signals across reviewed studies for anxiety detection.** **a** Unique features by signal: Number of unique features extracted from each signal modality (ECG, RSP, EDA, and PPG). **b** Most used ECG features: ECG features reported more than once across studies, highlighting HRV features (e.g., Mean HR, pNN50, and SDNN). **c** Most used RSP features: RSP features reported more than once across studies, focusing on respiratory timing and interval measures (e.g., Ti/Te mean, Mean IBI). **d** Most used EDA features: EDA features reported more than once across studies, highlighting amplitude and peak-related metrics (e.g., Mean EDA, Mean Peak Amp SCR). **e** Most used PPG features: PPG features reported more than once across studies, emphasizing heart rate and variability metrics (e.g., Mean HR, SD PPG). Note: In (a), *n* indicates the number of unique features identified across studies for each signal modality. In (b–e), *n* indicates the number of independent studies reporting the use of each specific feature. ECG features: Mean HR mean heart rate, pNN50 percentage of successive RR intervals differing by more than 50 ms, SDNN standard deviation of RR intervals, RMSSD root mean square of successive RR interval differences, HF power high-frequency power of heart rate variability, LF power low-frequency power of heart rate variability, Mean RR mean R-R interval, Mean aFDnRR mean of the absolute values of the first differences of normalized RR intervals, Mean FDRR mean of the first differences of RR intervals, SD aFDRR standard deviation of the absolute values of the first differences of RR intervals, VLF power very low-frequency power of RR intervals, LF norm normalized low-frequency power, LF/HF ratio of

low-frequency to high-frequency power, SD HR standard deviation of heart rate, Var ECG variance of the ECG signal, HF norm normalized high-frequency power, TINN triangular interpolation of NN interval histogram, SD ECG standard deviation of the ECG signal, Median ECG median of the ECG signal, Mean ECG mean of the ECG signal, Max ECG maximum amplitude of the ECG signal, Min ECG minimum amplitude of the ECG signal. PPG features: mean HR mean heart rate, SD PPG standard deviation of the PPG signal, Mean PPI mean pulse interval, Mean PPG mean of the PPG signal, LF/HF ratio of low-frequency to high-frequency power. EDA features: mean EDA mean electrodermal activity amplitude, Mean Peak Amp SCR mean amplitude of skin conductance response peaks, SD EDA standard deviation of the EDA signal amplitude, Min EDA minimum amplitude of the EDA signal, Max Peak Amp SCR maximum amplitude of SCR peaks, Max EDA maximum amplitude of the EDA signal, SCR Rate rate of skin conductance response peaks per unit time, Median EDA median of the EDA signal amplitude, Var EDA variance of the EDA signal amplitude, Mean SCL mean of the tonic component (skin conductance level) of the EDA signal, Median Peak Amp SCR median of the amplitudes of detected peaks in the SCR. RSP features: Ti/Te mean mean inspiratory to expiratory time ratio, Mean IBI mean inter-breath interval, SD ReR standard deviation of the respiratory rate, Mean ReR mean respiratory rate, Te mean mean expiratory time, Ti mean mean inspiratory time, SD of IBIs standard deviation of inter-breath intervals.

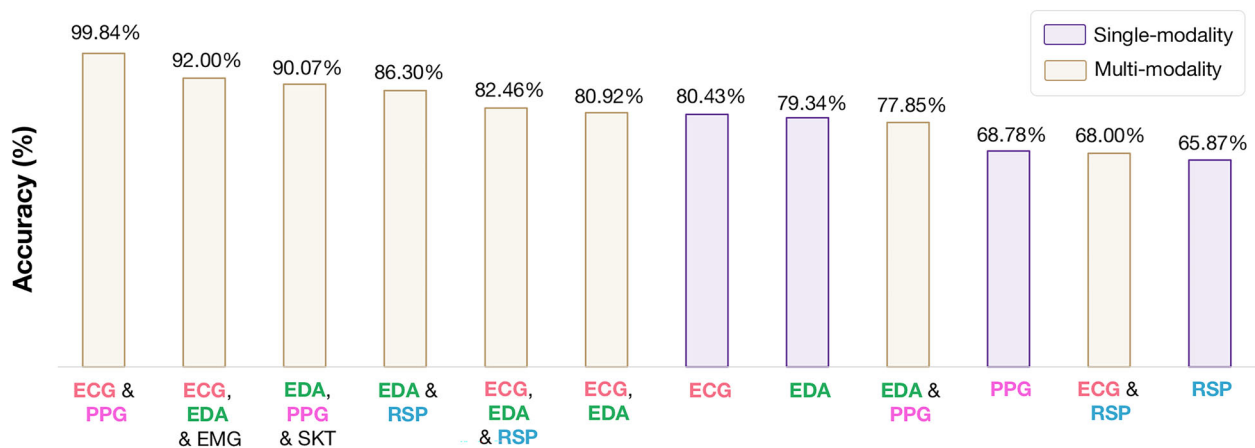
which provided F1 scores, but did not report accuracy, were excluded from our calculations. Among single-modality results, ECG exhibited the highest pooled accuracy at 80.43%, with a mean accuracy of 81.10% ( $\pm 14.70\%$ ), based on results from 12 different studies<sup>30–38,44,47,50</sup>, with accuracies ranging from 57.65<sup>37</sup> to 99.95%<sup>30</sup>. EDA followed with a pooled accuracy of 79.34% and a mean accuracy of 76.92% ( $\pm 17.46\%$ ), based on results from six studies<sup>18–21,44,47</sup>, with accuracies ranging from 42.53<sup>18</sup> to 89.00%<sup>19</sup>. PPG signals demonstrated a pooled accuracy of 68.78%, with a mean accuracy of 66.83% ( $\pm 10.94\%$ ), based on five studies<sup>18–22,50</sup>, ranging from 49.75<sup>18</sup> to 78.00%<sup>19</sup>. RSP, though showing the lowest pooled accuracy at 65.87%, had a wider variation in results, with a mean accuracy of 76.40% ( $\pm 22.05\%$ ) from four studies<sup>39–41,47</sup>, ranging from 53.00<sup>47</sup> to 97.90%<sup>40</sup>. Overall, single-modality approaches revealed a pooled accuracy of 76.98% and a mean accuracy of 76.83% ( $\pm 15.88\%$ ) across 27 reported results from 17 different studies.

In studies with multi-modality approaches, ECG & PPG achieved the highest accuracy of 99.84%, although it was based on a single study<sup>50</sup>. ECG, EDA & EMG achieved an accuracy of 92.00%, also based on a single article<sup>49</sup>. The combination of EDA, PPG & SKT demonstrated a pooled accuracy of 90.07% and a mean accuracy of 92.89% ( $\pm 9.32\%$ ), based on two studies, 86.30<sup>20</sup> and 99.48%<sup>21</sup>, respectively. The combination of EDA & RSP yielded an accuracy of 88.00%, also derived from a single study<sup>47</sup>. ECG, EDA & RSP achieved a pooled accuracy of 82.46% and a mean accuracy of 82.50% ( $\pm 3.54\%$ ), based on two studies, achieving 80.0%<sup>48</sup> and 85.00%<sup>47</sup>, respectively. ECG & EDA, a commonly used combination, reported a pooled accuracy of 80.92% and a mean accuracy of 85.60% ( $\pm 14.35\%$ ) across five results from four studies<sup>44–47</sup>. Accuracies ranged from 64.50 to 99.00%, both provided by Zhou et al.<sup>45</sup>, who trained their model on two distinct datasets, APD<sup>57</sup> (64.50%) and WESAD<sup>51</sup> (99.00%). This wide range illustrates how

### a) Accuracy results across studies by signal(s) used



### b) Pooled accuracies by signal(s) used



**Fig. 6 | Accuracy results for anxiety detection across studies by signal(s) used.**

**a** Accuracy results by signal(s) used: Boxplots illustrate the distribution of reported accuracies across studies, where the centre line indicates the median, the box represents the interquartile range (IQR), and whiskers show the minimum and maximum values. Categories with fewer than three studies are shown as individual points without boxplots. Single-modality results are highlighted in green, while multi-modality results are shown in purple. **b** Pooled accuracies for anxiety

detection: Bar chart displaying pooled accuracy values for each signal modality or combination, calculated as weighted averages based on the number of studies reporting each result. **Note:** In (a), *n* indicates the number of independent study results contributing to each boxplot. ECG electrocardiogram, EDA electrodermal activity, PPG photoplethysmography, RSP respiration, SKT skin temperature, EMG electromyography.

dataset-specific factors, such as the number of participants (52 for APD versus 15 for WESAD) and experimental protocols (public speaking and bug-box tasks for APD versus TSST for WESAD), can significantly influence model performance. EDA & PPG achieved a pooled accuracy of 77.85% and a mean accuracy of 78.95% ( $\pm 25.75\%$ ) based on three studies<sup>18,42,43</sup>, ranging from 49.26<sup>18</sup> to 95.22%<sup>42</sup>. Lastly, the combination of ECG & RSP yielded an accuracy of 68.00% based on a single study<sup>47</sup>. Multi-modality approaches collectively achieved a pooled accuracy of 81.98% and a mean accuracy of 85.22% ( $\pm 14.29\%$ ) across 16 reported results from 12 different studies.

#### Assessment of study limitations

A systematic assessment of 26 studies revealed varying levels of limitations across eight predefined categories (Fig. 7). The most prevalent serious limitations were found in population representation (L1) (38%),

sample size (L2) (38%), and reproducibility (L3) (58%), reflecting challenges in ensuring diverse, generalizable findings as well as availability of datasets and code for replication. In contrast, study completion rate (L3) and algorithmic transparency (L6) were frequently rated as Low (81% and 85%, respectively), suggesting consistent reporting of dropout rates and well-explained methods. Selective reporting (L4) and validation robustness (L7) were primarily rated moderate (77% and 65%, respectively), highlighting common gaps in reporting secondary evaluation metrics, such as sensitivity, specificity, and F1 scores. Furthermore, while many studies relied on standard validation techniques, such as 5-fold or 10-fold cross-validation or hold-out validation, more rigorous methods, including LOSO CV or validation on independent datasets, were frequently absent. Outcome measurement validity (L5) often relied on partially validated methods, such as self-reports (54%), raising concerns about reliability.

**Table 1 | Summary of pooled and mean accuracies across signal modalities for anxiety detection**

| Signal                 | Pooled accuracy (%) | Mean accuracy* (%)          | Number of results |
|------------------------|---------------------|-----------------------------|-------------------|
| <b>Single-modality</b> |                     |                             |                   |
| ECG                    | 80.43               | 81.10 ± 14.70 (57.65–99.95) | 12                |
| EDA                    | 79.34               | 76.92 ± 17.46 (42.53–89.00) | 6                 |
| PPG                    | 68.78               | 66.83 ± 10.94 (49.75–78.00) | 5                 |
| RSP                    | 65.87               | 76.40 ± 22.05 (53.00–97.90) | 4                 |
| Single-modality        | 76.98               | 76.83 ± 15.88 (42.53–99.95) | 27                |
| <b>Multi-modality</b>  |                     |                             |                   |
| ECG & PPG              | 99.84               | 99.84                       | 1                 |
| ECG, EDA & EMG         | 92.00               | 92.00                       | 1                 |
| EDA, PPG & SKT         | 90.07               | 92.89 ± 9.32 (86.30–9.48)   | 2                 |
| EDA & RSP              | 88.00               | 88.00                       | 1                 |
| ECG, EDA & RSP         | 82.46               | 82.50 ± 3.54 (80.00–85.00)  | 2                 |
| ECG & EDA              | 80.92               | 85.60 ± 14.35 (64.50–99.00) | 5                 |
| EDA & PPG              | 77.85               | 78.95 ± 25.75 (49.26–95.22) | 3                 |
| ECG & RSP              | 68.00               | 68.00                       | 1                 |
| Multi-modality         | 81.98               | 85.22 ± 14.29 (49.26–99.84) | 16                |

\*Mean accuracy is reported as mean ± standard deviation (minimum–maximum). ECG electrocardiogram, EDA electrodermal activity, PPG photoplethysmography, RSP respiratory signal.

## Discussion

We conducted a comprehensive review of the emerging evidence regarding the use of physiological signals, including ECG, respiratory signals, EDA, and PPG, to detect anxiety. Most studies selected for this review primarily focused on affective or emotional states rather than clinical populations with medically diagnosed anxiety. Though this growing research area is still in its early stages, the findings from this review may offer preliminary insights into its clinical impact and lay the groundwork for subsequent studies. Below, we discuss the implications of our findings and potential applications.

Comprising of the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS), the autonomic nervous system (ANS) plays a pivotal role in connecting physiological signals to anxiety. The ANS comprises the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). When an individual experiences anxiety, the typical response involves the activation of the SNS and the inhibition of the PNS<sup>66</sup>. Slow breathing can potentially enhance the activation of the PNS<sup>67</sup>. Both sympathetic and parasympathetic activities can influence the spontaneous sinus node depolarization, resulting in alterations in cardiac rate and rhythm<sup>68</sup>. For instance, HRV metrics, which can be extracted from ECG and PPG signals, provide insights into how the balance between the SNS and PNS influences cardiac activity<sup>69</sup>.

Several studies<sup>70,71</sup> reviewed here have indicated that the HRV extracted from the ECG diminishes when the subject feels anxious. Decreased time-domain HRV metrics, such as RMSSD<sup>33,37,38,44,45,47,50</sup>, SDNN<sup>35,37,38,44,45,47,50</sup>, and pNN50<sup>35,37,38,47,49,50</sup> have consistently been used in studies to detect anxiety, demonstrating their effectiveness in identifying anxiety-induced reductions in parasympathetic activity. These features are similarly robust when extracted from PPG signals, with observed reductions in RMSSD<sup>20,22,50</sup> and SDNN<sup>20,22,50</sup> reported during anxiety states. In addition to HRV, HR itself has proven to be a reliable marker of anxiety, with increases in HR commonly reported during anxiety episodes. This trend is observed in both ECG-derived measures<sup>36,43–47,50</sup> and PPG-derived measures<sup>19,20,22,42</sup>. Frequency-domain features, such as LF power and HF power<sup>34,35,37,38,45</sup>, and VLF power<sup>34,37,38</sup>, extracted from ECG, along with the LF/HF ratio—an indicator of autonomic shifts reflecting increased sympathetic dominance in anxiety states—extracted from both ECG<sup>37,38,45,47</sup> and PPG<sup>18,22</sup>, are crucial for understanding the balance between sympathetic and parasympathetic activity during anxiety. Together, these features provide a detailed understanding of autonomic changes associated with anxiety. The consistent

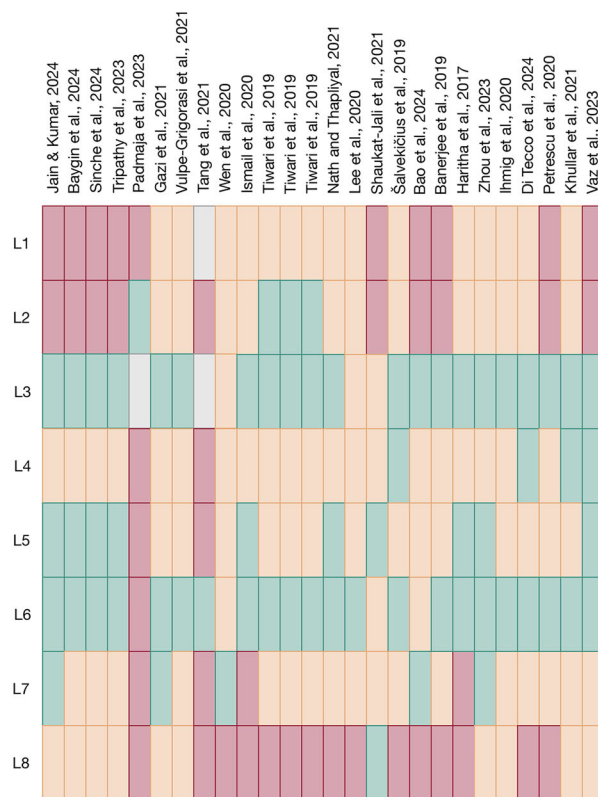
findings across studies of both ECG and PPG signals highlight the reliability of time- and frequency-domain HRV metrics for accurate anxiety detection.

Complementing HRV metrics, EDA provides a direct measure of SNS activity, making it a valuable signal for anxiety detection, as well. Unlike HRV, which reflects the balance between SNS and PNS, EDA specifically captures sympathetic arousal associated with heightened anxiety states. Commonly used features, such as mean EDA<sup>18,20,44–46</sup>, peak SCR amplitude<sup>19,20,44,47</sup>, and SD EDA<sup>18,20,44,45</sup>, capture both baseline arousal and response intensity with high sensitivity to anxiety-induced SNS changes.

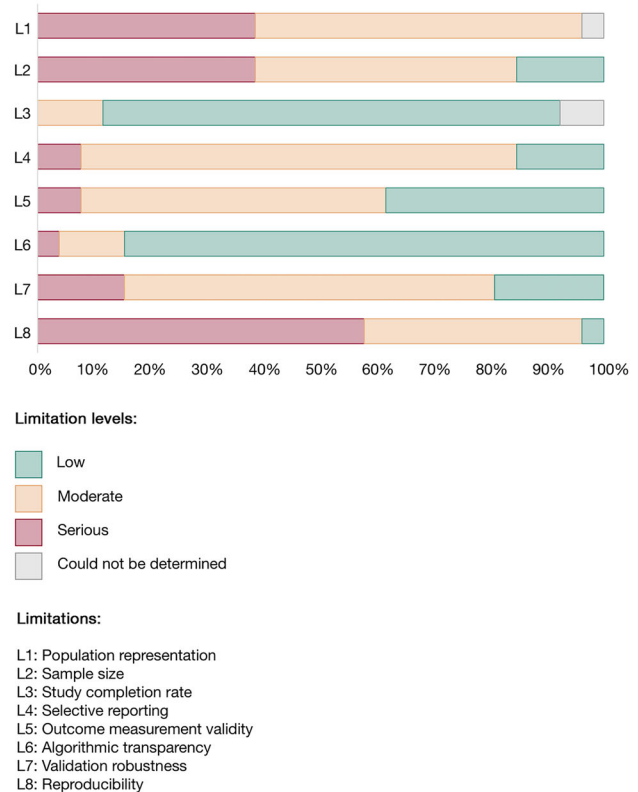
ADs can also lead to respiratory changes. Ihmig et al.<sup>46</sup> reported an increase in respiration rates during anxiety. Haritha et al.<sup>41</sup> suggested that breathing patterns could provide ample information for anxiety detection, but they did not specify how these patterns changed with anxiety. Features such as the mean inspiratory to expiratory time ratio<sup>39,40,47</sup>, mean and standard deviation of respiratory rate, as well as mean interbreath intervals<sup>39,41,47</sup>, which were among the most commonly used respiratory features in this review, could be promising for detecting anxiety states.

The reviewed studies show that multi-modality approaches generally outperform single-modality methods in terms of accuracy (81.94 vs. 76.85%), likely due to the integration of complementary signals. However, fewer studies reported multi-modality results (16 vs. 27 for single-modality), and most combinations were based on only one or a few studies. The findings should be interpreted with caution, as the limited number of studies and variability in methods, such as different sample sizes, anxiety induction and assessment tools, as well as data preprocessing and machine learning methods, preclude conclusive comparisons of signal efficacy. Signals reflecting ANS activity, such as ECG, PPG, and EDA, demonstrated higher accuracies as compared to RSP, which indirectly reflects ANS activity. Interestingly, integrating RSP with ANS-related signals, such as EDA & RSP (88.00%, based on one study<sup>47</sup>) or ECG, EDA & RSP (82.46%, based on two studies<sup>47,48</sup>), achieved pooled accuracies higher than EDA alone, suggesting that RSP may enhance detection performance. However, RSP results were based on four studies only and exhibited the widest accuracy range (53.00–97.90%), indicating significant methodological differences and limited representation. Among single modalities, ECG yielded the most reliable results, supported by the highest number of studies and a pooled accuracy of 80.42%. Furthermore, many of the highest-accuracy results stem from basic early-fusion approaches (e.g., simple feature concatenation) with minimal feature selection or validation<sup>21,22,44–46,49,50</sup>. In contrast, studies employing structured feature selection or decision-level fusion (e.g.,<sup>42,43,47,48</sup>)

## a) Limitation levels across studies



## b) Distribution of limitation levels by category



**Fig. 7 | Limitation levels across included studies and their distribution by category.** **a** Limitation levels across studies: Heatmap summarizing the limitation levels for population representation, sample size, study completion rate, selective reporting, outcome measurement validity, algorithmic transparency, validation robustness, and reproducibility (L1–L8) for each study. Levels are categorized as Low

(green), Moderate (orange), Serious (red), or Could Not Be Determined (gray). **b** Distribution of limitation levels by category: Bar chart illustrating the proportion of studies rated as Low, Moderate, Serious, or Could Not Be Determined for each limitation category (L1–L8), highlighting the prevalence of Serious limitations in population representation (L1) and sample size (L2). **Note:** L Limitation.

reported more moderate yet likely generalizable results. These studies applied methods such as feature permutation importance<sup>42,47</sup>, chi-squared tests<sup>42</sup>, sequential feature selection (SFS)<sup>42,46</sup>, and regression-based ranking to refine models and mitigate redundancy. This pattern suggests that some of the highest-reported accuracies may be inflated by overfitting rather than true multimodal complementarity, emphasizing the need for rigorous feature selection. Given the limited evidence for both multi- and single-modality approaches, further research with standardized protocols and larger datasets is essential. Despite these limitations, leveraging ANS-related signals and integrating diverse modalities remain promising strategies for improving anxiety detection.

We also observed that technology can discern anxiety across various stimuli and experimental paradigms, including public speaking tasks<sup>19–21,35,44,45,49</sup>, virtual reality-based exposure to phobias such as spiders and heights<sup>33,43,46–48</sup>, video-based anxiety induction using general anxiety-inducing clips, horror trailers, and driving scenarios<sup>18,30,31,42</sup>, workplace stress monitoring<sup>37–39</sup>, and tasks involving relaxation, gaming, or social anxiety<sup>22,36,40</sup>. These findings highlight the adaptability of physiological signal-based methods to a wide range of anxiety-inducing challenges, supporting their potential utility in diverse real-world scenarios.

While we cannot advocate for ECG-, EDA-, PPG- and RSP-based wearable technologies as the sole method for detecting anxiety, their value and potential utility in clinical practice as a supportive tool is clear and only expected to grow along with the advancement and enhanced robustness of these technologies. Unlike traditional 12-lead ECGs, which are typically used in clinical settings and require professional administration, WDs can be used continuously in naturalistic environments. This makes them

especially valuable as a preliminary screening tools or triage mechanisms for patients in rural areas or developing countries who face significant access barriers to prompt medical and mental health care. By providing an early indication of potential anxiety, they can facilitate timely referrals and access to necessary care<sup>72,73</sup>. Additionally, wearable technologies are beneficial for individuals with Autism Spectrum Disorder or others who face challenges with communication and self-reflection<sup>74</sup>. Such individuals may find traditional self-reporting scales difficult, making WDs an attractive alternative for objectively detecting anxiety and ensuring more accessible and timely treatment. Although current research is still exploring the capability of these technologies to distinguish anxiety from other mental health disorders with overlapping symptoms, initial findings on PPG have identified distinctions in HRV between patients with anxiety and depression<sup>75,76</sup>. Nevertheless, further multidisciplinary investigation is essential to determine the specific sensitivity and accuracy of wearables in this context and identify discriminating features. For clinicians in a rapidly emerging precision medicine world, these technologies are poised to be integrated in traditional diagnostic and monitoring protocols of patients diagnosed with social phobia, psychological stress, or anxiety. Wearable technologies provide a dynamic view of patients' progress, enabling for real-time therapeutic adjustments and effectiveness assessments, complementing traditional symptom evaluations during in-person consultations. Importantly, considering the high prevalence of AD in adolescents throughout the world, providing technology-driven tools for assessment, monitoring and treatment is invaluable for a generation that is both technology reliant and tech-savvy.

Early mHealth research on mental health primarily focused on text messaging interventions for psychological support. Studies have shown that

text messaging, or short message service (SMS), is effective for addressing mental health issues, including anxiety<sup>77</sup>. Moreover, research in telemental health has been centered around creating accessible virtual mental health consultations for patients. Such studies indicate that telemental health effectively addresses mental health challenges and enhances patient accessibility<sup>78</sup>. Given the advancements in wearable technologies, there's potential for a more integrated digital healthcare approach. Specifically, wearable signals could be incorporated into a wide array of digital mental health tools. This technology might prove valuable in tandem with telemental health consultations, as it offers objective physiological data in addition to subjective self-reported symptoms of anxiety or social phobias. This could provide clinicians with a more comprehensive, informed and data-driven clinical overview. However, more research is needed to validate this technology before endorsing it as an additional tool and rigorous regulatory measures should be applied towards identifying and approving legitimate interventions.

This review highlights significant advancements in the field of physiological signal-based anxiety detection but is not without limitations. The scarcity of available studies - most multi-modality results were based on single studies, and all single-signal modalities except ECG had a total of six or fewer studies - limits the generalizability of our findings. Results from both single- and multi-modality studies showed a wide range of accuracies, emphasizing study-specific variability. For instance, Zhou et al. reported accuracies of 64.50 and 99.00% for ECG and EDA signals when tested on the APD and WESAD datasets, respectively, highlighting the influence of factors such as sample size, anxiety induction methods, and measurement tools on model performance. Many of the reviewed studies relied on small sample sizes and lacked standardization in anxiety induction methods, feature extraction, and validation approaches. These inconsistencies make it challenging to draw definitive conclusions about the reliability of these technologies across diverse settings. While the reviewed studies focused on controlled anxiety induction in healthy participants, validating wearables for anxiety screening requires both cross-sectional studies comparing diagnosed anxiety patients to healthy controls using standardized tools like Generalized Anxiety Disorder Assessment scale<sup>79</sup>, and longitudinal studies assessing whether wearables can detect emerging anxiety disorders. Without these validations, their ability to differentiate temporary stress from chronic anxiety remains uncertain. Gold-standard randomized controlled trials remain essential for establishing clinical validity<sup>80</sup>. Further research should also explore their efficacy in detecting general phobias, psychological stress, and specific anxiety triggers. Moreover, it is crucial to acknowledge that certain medical conditions, such as cardiovascular diseases, can influence ECG and PPG readings. This confounding factor might compromise the precision of anxiety detection through this method<sup>81</sup>. Hence, when executing ECG and PPG analysis for anxiety detection, researchers should tread with caution. Participants with medical issues that might skew the accuracy of ECG and PPG signals should either be omitted from the study, or their conditions should be factored into the data analysis.

With a growing number of people affected by ADs all over the world, there is a pressing need for a dependable anxiety detection system suitable for home use. Gazi et al.<sup>47</sup> pointed out the potential challenges in generalizing the model to real-world settings, which often entail multiple stimuli differing from controlled laboratory conditions. Individuals diagnosed with specific mental disorders might react differently to identical stimuli compared to their healthy counterparts. Expert clinicians have highlighted that while healthy individuals typically exhibit moderate breath-to-breath variability, those with ADs tend to show reduced variability in the same parameter<sup>82,83</sup>. Some research<sup>84,85</sup> has shown that the HRV of individuals with mental disorders (e.g., depression and anxiety) is lower compared to the general healthy populace. However, only one study<sup>41</sup> encompassed data from both healthy participants and those diagnosed with ADs. Based on our review, several recommendations can guide the development of more reliable, technology-driven AD assessment. Future studies should collect

biosignals from larger and more diverse populations, with detailed information on age, gender, health status, and ethnicity. Data should capture both calm and anxious states from the same individuals, across a range of mental health states, and ideally in real-world rather than only controlled laboratory settings. Synchronized multimodal signals such as ECG, EDA, PPG, and RSP should be prioritized to better capture the multifactorial markers of anxiety. Additionally, usability issues should be addressed, prioritizing smaller, user-friendly designs that minimize distraction during daily activities. Finally, advanced classifiers should be leveraged to move beyond simple two-level detection, and models should be evaluated using a broader set of metrics, including MCC, ACC, F1-score, sensitivity, and specificity, to enable comprehensive and comparable performance assessments across studies.

The suggestions put forth can prove instrumental in creating reliable and varied datasets. Such datasets can enable researchers to train models with enhanced accuracy for anxiety detection. By sharing these datasets and generalizing their findings across different populations, the robustness and accuracy of machine learning models in real-world scenarios can be improved. Ultimately, these datasets may pave the way for more efficient mental health assessments. The advent of smaller WDs allows for continuous monitoring of bio-signals during an individual's daily activities, promoting personalized mental health detection and timely interventions. Leveraging advanced machine learning algorithms holds promise in refining the accuracy of anxiety detection. Adopting a consistent evaluation metric would simplify comparisons across various studies and methodologies in the realm of anxiety detection.

## Conclusions

This study aimed to assess the efficacy of ECG, EDA, PPG and RSP signals, gathered through various WDs, in the detection of anxiety, while also summarizing the types of WDs used and the features extracted from the different signals. One constraint of our study is that based on the inclusion/exclusion criteria, it only encompassed 26 studies, potentially limiting the broader applicability of our findings. ECG emerged as the most reliable single modality, with robust performance across studies, underscoring its reliability in monitoring psychophysiological changes associated with anxiety states. RSP showed the widest variability and lowest pooled accuracy, underscoring the need for methodological refinement. Although multimodal approaches generally outperformed single-modality methods, the limited number of reviewed studies indicates the need for further research. Most reviewed studies used small datasets, focused on healthy participants, and lacked standardized protocols, limiting the generalizability of the results. Addressing these issues, such as variability in anxiety induction methods, sample sizes, and evaluation metrics, is crucial for achieving comparable outcomes. Considering that episodes of anxiety can manifest unpredictably, it is crucial for individuals to possess a portable WD that offers real-time anxiety evaluation. However, a majority of the studies we reviewed relied on pre-existing databases and were restricted to retrospective signal analysis. Early identification of functional impairments during daily tasks is paramount for prompt clinical intervention. Wearable technologies have immense potential as complementary tools in mental health care, facilitating continuous monitoring, early detection, and timely intervention. By addressing current limitations and focusing on robust datasets and standardized methodologies, research can advance the development of reliable, real-time, technology-driven anxiety detection systems.

## Data availability

This study is a systematic review and did not generate any new datasets. All data supporting the findings are derived from previously published studies, which are cited in the article and detailed in Supplementary Data 1 and Supplementary Data 2.

Received: 21 February 2025; Accepted: 30 October 2025;

Published online: 09 January 2026

## References

- World Health Organization. Depression and other common mental disorders: global health estimates. Technical Report, World Health Organization, Geneva (WHO, 2017).
- World Health Organization. Mental health of adolescents <https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health> (2024). Accessed: 10 February 2025.
- Vos, T. et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204–1222 (2020).
- Beck, A. T., Emery, G. & Greenberg, R. L. *Anxiety Disorders and Phobias: A Cognitive Perspective*. (Basic Books, 2005).
- Tully, P. J., Harrison, N. J., Cheung, P. & Cosh, S. Anxiety and cardiovascular disease risk: a review. *Curr. Cardiol. Rep.* **18**, 1–8 (2016).
- Cohen, B. E., Edmondson, D. & Kronish, I. M. State of the art review: depression, stress, anxiety, and cardiovascular disease. *Am. J. Hypertension* **28**, 1295–1302 (2015).
- Holt, R. I. et al. The relationship between depression, anxiety and cardiovascular disease: findings from the Hertfordshire Cohort Study. *J. Affect. Disord.* **150**, 84–90 (2013).
- Davies, S. J. & Allgulander, C. Anxiety and cardiovascular disease. *Anxiety Disord.* **29**, 85–97 (2013).
- Cheung, R., O'Donnell, S., Madi, N. & Goldner, E. Factors associated with delayed diagnosis of mood and/or anxiety disorders. *Health Promotion Chronic Dis. Prev. Can. Res. policy Pract.* **37**, 137 (2017).
- Perrotta, G. Anxiety disorders: definitions, contexts, neural correlates and strategic therapy. *J. Neur. Neurosci.* **6**, 042 (2019).
- Ates, H. C. et al. End-to-end design of wearable sensors. *Nat. Rev. Mater.* **7**, 887–907 (2022).
- Lyzwinski, L. N., Elgendi, M. & Menon, C. The use of photoplethysmography in the assessment of mental health: Scoping review. *JMIR Ment. Health* **10**, e40163 (2023).
- Lu, G., Yang, F., Taylor, J. A. & Stein, J. F. A comparison of photoplethysmography and ecg recording to analyse heart rate variability in healthy subjects. *J. Med. Eng. Technol.* **33**, 634–641 (2009).
- Lu, G. & Yang, F. Limitations of oximetry to measure heart rate variability measures. *Cardiovasc. Eng.* **9**, 119–125 (2009).
- Elgendi, M. & Menon, C. Assessing anxiety disorders using wearable devices: challenges and future directions. *Brain Sci.* **9**, 50 (2019).
- Giardino, N. D., Friedman, S. D. & Dager, S. R. Anxiety, respiration, and cerebral blood flow: implications for functional brain imaging. *Compr. Psychiatry* **48**, 103–112 (2007).
- Elgendi, M. & Menon, C. Machine learning ranks ECG as an optimal wearable biosignal for assessing driving stress. *IEEE Access* **8**, 34362–34374 (2020).
- Lee, S., Lee, T., Yang, T., Yoon, C. & Kim, S.-P. Detection of drivers' anxiety invoked by driving situations using multimodal biosignals. *Processes* **8**, <https://www.mdpi.com/2227-9717/8/2/155> (2020).
- Nath, R. & Thapliyal, H. Machine learning-based anxiety detection in older adults using wristband sensors and context feature. *SN Comp. Sci.* **2**, 359 (2021).
- Šalkevičius, J., Damaševičius, R., Maskeliūnas, R. & Laukienė, I. Anxiety level recognition for virtual reality therapy system using physiological signals. *Electronics* **8**, <https://www.mdpi.com/2079-9292/8/9/1039> (2019).
- Shaukat-Jali, R., van Zalk, N. & Boyle, D. Detecting subclinical social anxiety using physiological data from a wrist-worn wearable: Small-scale feasibility study. *JMIR Formative Res.* **5**, e32656 (2021).
- Bao, Y., Xue, M. & Gohumpu, J. et al. Prenatal anxiety recognition model integrating multimodal physiological signal. *Sci. Rep.* **14**, 21767 (2024).
- Hickey, B. et al. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review. *Sensors* **21**, 3461 (2021).
- Shikha, S., Sethia, D. & Indu, S. A systematic review on physiology-based anxiety detection using machine learning. *Biomed. Phys. Eng. Express* (2025).
- Sajno, E. et al. Machine learning in biosignals processing for mental health: A narrative review. *Frontiers in Psychology* **Volume 13 - 2022**, <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.1066317> (2023).
- Ahmed, A. et al. Wearable devices for anxiety & depression: a scoping review. *Computer Methods Prog. Biomedicine Update* **3**, 100095 (2023).
- Liberati, A. et al. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J. Clin. Epidemiol.* **62**, e1–e34 (2009).
- Deeks, J. J., Higgins, J. P., Altman, D. G. & Group, C. S. M. Analysing data and undertaking meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions* 241–284 (Wiley, 2019).
- Bird, K. et al. Assessment of hypertension using clinical electrocardiogram features: a first-ever review. *Front. Med.* **7**, 583331 (2020).
- Baygin, M. et al. Automated anxiety detection using probabilistic binary pattern with ECG signals. *Comp. Methods Prog. Biomed.* **247**, 108076 (2024).
- Tripathy, R. K., Dash, D. K., Ghosh, S. K. & Pachori, R. B. Detection of different stages of anxiety from single-channel wearable ECG sensor signal using Fourier-Bessel domain adaptive wavelet transform. *IEEE Sens. Lett.* **7**, 1–4 (2023).
- Padmaja, K. V. et al. Anxiogram: Unmasking anxiety with IoT-enhanced ECG. In *2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, 1–4 (CSITSS, 2023).
- Vulpe-Grigorași, A. & Grigore, O. A neural network approach for anxiety detection based on ECG. In *Proc. International Conference on e-Health and Bioengineering (EHB)*, 1–4 (EHB, 2021).
- Tang, Y., Zhou, N. & Guan, A. Anxiety detection algorithm for wearable devices based on DBSCAN. In *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, 56–60 (2021).
- Wen, W. et al. Toward constructing a real-time social anxiety evaluation system: exploring effective heart rate features. *IEEE Trans. Affect. Comput.* **11**, 100–110 (2020).
- Ismail, N. M., Airij, A. G., Sudirman, R. & Omar, C. Early detection of social anxiety disorder by using screening tools and wearable sensors. In *Proc. 6th International Conference on Computing Engineering and Design (ICCED)*, 1–6 (ICCED, 2020).
- Tiwari, A., Cassani, R., Narayanan, S. & Falk, T. H. A comparative study of stress and anxiety estimation in ecological settings using a smart-shirt and a smart-bracelet. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2213–2216 (EMBC, 2019).
- Tiwari, A., Narayanan, S. & Falk, T. H. Stress and anxiety measurement “in-the-wild” using quality-aware multi-scale HRV features. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 7056–7059 (EMBC, 2019).
- Tiwari, A., Narayanan, S. & Falk, T. H. Breathing rate complexity features for “in-the-wild” stress and anxiety measurement. In *2019 27th European Signal Processing Conference (EUSIPCO)*, 1–5 (EUSIPCO, 2019).
- Banerjee, T., Khasnobish, A., Chowdhury, A. & Chatterjee, D. Reckoning respiratory signals to affectively decipher mental state. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4654–4659 (EMBC, 2019).
- Jariritha, H., Negi, S., Menon, R. S., Kumar, A. A. & Kumar, C. S. Automating anxiety detection using respiratory signal analysis. In *2017 IEEE Region 10 Symposium (TENSYP)*, 1–5 (TENSYP, 2017).

42. Di Tecco, A., Pistolesi, F. & Lazzerini, B. Elicitation of anxiety without time pressure and its detection using physiological signals and artificial intelligence: a proof of concept. *IEEE Access* **12**, 22376–22393 (2024).
43. Petrescu, L. et al. Integrating biosignals measurement in virtual reality environments for anxiety detection. *Sensors* **20**, 7088 (2020).
44. Jain, A. & Kumar, R. Machine learning based anxiety detection using physiological signals and context features. In *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, 116–121 (ICACCT, 2024).
45. Zhou, E., Soleymani, M. & Matarić, M. J. Investigating the generalizability of physiological characteristics of anxiety. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 4848–4855 (2023).
46. Ihmig, F. R., Neurohr-Parakenings, F., Schäfer, S. K., Lass-Hennemann, J. & Michael, T. On-line anxiety level detection from biosignals: machine learning based on a randomized controlled trial with spider-fearful individuals. *PLoS One* **15**, e0231517 (2020).
47. Gazi, A. H. et al. Respiratory markers significantly enhance anxiety detection using multimodal physiological sensing. In *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, 1–4 (IEEE, 2021).
48. Khullar, V., Tiwari, R. G., Agarwal, A. K. & Dutta, S. Physiological signals based anxiety detection using ensemble machine learning. In Tavares, J. M. R. S., Dutta, P., Dutta, S. & Samanta, D. (eds.) *Cyber Intelligence and Information Retrieval*, 597–608 (Springer, 2022).
49. Vaz, M., Summavielle, T., Sebastião, R. & Ribeiro, R. P. Multimodal classification of anxiety based on physiological signals. *Appl. Sci.* **13**, <https://www.mdpi.com/2076-3417/13/11/6368> (2023).
50. Sinche, S., Acán, J. & Hidalgo, P. Anxiety detection using consumer heart rate sensors. *Eng. Proc.* **77**, <https://www.mdpi.com/2673-4591/77/1/10> (2024).
51. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C. & Laerhoven, K. V. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 2018 International Conference on Multimodal Interaction (ICMI)* (ACM, 2018).
52. Kirschbaum, C., Pirke, K.-M. & Hellhammer, D. H. The 'trier social stress test'—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* **28**, 76–81 (1993).
53. Marteau, T. M. & Bekker, H. The development of a six-item short-form of the state scale of the Spielberger state-trait anxiety inventory (STAI). *Br. J. Clin. Psychol.* **31**, 301–306 (1992).
54. Ihmig, F. R., Gogeaşcoachea, A., Schäfer, S., Lass-Hennemann, J. & Michael, T. Electrocardiogram, skin conductance and respiration from spider-fearful individuals watching spider video clips (version 1.0.0) <https://doi.org/10.13026/sq6q-zg04> (2020).
55. Elgendi, M., Galli, V., Ahmadzadeh, C. & Menon, C. Dataset of psychological scales and physiological signals collected for anxiety assessment using a portable device. *Data* **7**, <https://www.mdpi.com/2306-5729/7/9/132> (2022).
56. Schneider, H., Esbitt, S. & Gonzalez, J. Hamilton anxiety rating scale. In Gellman, M. & Turner, J. (eds.) *Encyclopedia of Behavioral Medicine*, 924–925 [https://doi.org/10.1007/978-1-4419-1005-9\\_197](https://doi.org/10.1007/978-1-4419-1005-9_197) (Springer, 2013).
57. Senaratne, H., Kuhlmann, L., Ellis, K., Melvin, G. & Oviatt, S. A multimodal dataset and evaluation for feature estimators of temporal phases of anxiety. In *Proceedings of the 29th ACM International Conference on Multimedia* <https://doi.org/10.1145/3462244.3479900> (ACM, 2021).
58. Wolpe, J. Subjective units of distress scale (suds) <https://doi.org/10.1037/t05183-000> (1969).
59. Reuderink, B., Nijholt, A. & Poel, M. Affective Pacman: A frustrating game for brain-computer interface experiments. In *International Conference on Intelligent Technologies for Interactive Entertainment*, 221–227 (Springer, 2009).
60. Cassady, J. C. & Johnson, R. E. Cognitive test anxiety and academic performance. *Contemp. Educ. Psychol.* **27**, 270–295 (2002).
61. Furmark, T. et al. Social phobia in the general population: prevalence and sociodemographic profile. *Soc. Psychiatry Psychiatr. Epidemiol.* **34**, 416–424 (1999).
62. Liebowitz, M. R. Liebowitz social anxiety scale (LSAS) <https://doi.org/10.1037/t07671-000> (1987).
63. Association, A. P. *Diagnostic and Statistical Manual of Mental Disorders* <https://doi.org/10.1176/appi.books.9780890425596>. (American Psychiatric Publishing, Arlington, VA, 2013), 5th edn.
64. Gastinger, S., Donnelly, A., Dumond, R. & Prioux, J. A review of the evidence for the use of ventilation as a surrogate measure of energy expenditure. *J. Parenter. Enter. Nutr.* **38**, 926–938 (2014).
65. Chanel, G., Rebetez, C., Bétrancourt, M. & Pun, T. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era*, MindTrek '08, 13–17 (Association for Computing Machinery, 2008).
66. Ziegler, M. G. 50 - psychological stress and the autonomic nervous system. In Robertson, D., Biaggioni, I., Burnstock, G. & Low, P. A. (eds.) *Primer on the Autonomic Nervous System (Second Edition)*, 189–190 (Academic Press, 2004).
67. Jerath, R., Edry, J. W., Barnes, V. A. & Jerath, V. Physiology of long pranayamic breathing: Neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system. *Med. Hypotheses* **67**, 566–571 (2006).
68. Elghozi, J.-L., Girard, A. & Laude, D. Effects of drugs on the autonomic control of short-term heart rate variability. *Autonomic Neurosci.* **90**, 116–121 (2001).
69. Alberdi, A., Aztiria, A. & Basarab, A. Towards an automatic early stress recognition system for office environments based on multimodal measurements: a review. *J. Biomed. Inform.* **59**, 49–75 (2016).
70. Chalmers, J. A., Quintana, D. S., Abbott, M. J.-A. & Kemp, A. H. Anxiety disorders are associated with reduced heart rate variability: a meta-analysis. *Front. Psychiatry* **5**, 80 (2014).
71. Held, J., Vi^slä, A., Wolfer, C., Messerli-Bürgy, N. & Flückiger, C. Heart rate variability change during a stressful cognitive task in individuals with anxiety and control participants. *BMC Psychol.* **9**, 1–8 (2021).
72. Nielsen, M., D'Agostino, D. & Gregory, P. Addressing rural health challenges head on. *Mo. Med.* **114**, 363 (2017).
73. Morales, D. A., Barksdale, C. L. & Beckel-Mitchener, A. C. A call to action to address rural mental health disparities. *J. Clin. Transl. Sci.* **4**, 463–467 (2020).
74. Kushki, A., Khan, A., Brian, J. & Anagnostou, E. A Kalman filtering framework for physiological detection of anxiety-related arousal in children with autism spectrum disorder. *IEEE Trans. Biomed. Eng.* **62**, 990–1000 (2015).
75. Perpetuini, D. et al. Prediction of state anxiety by machine learning applied to photoplethysmography data. *PeerJ* **9**, e10448 (2021).
76. Liu, I., Ni, S. & Peng, K. Happiness at your fingertips: Assessing mental health with smartphone photoplethysmogram-based heart rate variability analysis. *Telemed. e-Health* **26**, 1483–1491 (2020).
77. Aguilera, A. et al. A text messaging intervention (staywell at home) to counteract depression and anxiety during covid-19 social distancing: pre-post study. *JMIR Ment. Health* **8**, e25298 (2021).
78. Langarizadeh, M. et al. Telemental health care, an effective alternative to conventional mental care: a systematic review. *Acta Inform. Med.* **25**, 240 (2017).
79. Löwe, B. et al. Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. *Med. Care* **46**, 266–274 (2008).
80. Hariton, E. & Locascio, J. J. Randomised controlled trials—the gold standard for effectiveness research. *BJOG Int. J. Obstet. Gynaecol.* **125**, 1716 (2018).

81. De Bacquer, D., De Backer, G., Kornitzer, M. & Blackburn, H. Prognostic value of ECG findings for total, cardiovascular disease, and coronary heart disease death in men and women. *Heart* **80**, 570–577 (1998).
82. Carney, R. M. et al. Depression, heart rate variability, and acute myocardial infarction. *Circulation* **104**, 2024–2028 (2001).
83. Choi, J. & Gutierrez-Osuna, R. Removal of respiratory influences from heart rate variability in stress monitoring. *IEEE Sens. J.* **11**, 2649–2656 (2011).
84. Kemp, A. H., Quintana, D. S., Felmingham, K. L., Matthews, S. & Jelinek, H. F. Depression, comorbid anxiety disorders, and heart rate variability in physically healthy, unmedicated patients: implications for cardiovascular risk. *PLoS one* **7**, e30777 (2012).
85. Cheng, Y.-C., Su, M.-I., Liu, C.-W., Huang, Y.-C. & Huang, W.-L. Heart rate variability in patients with anxiety disorders: a systematic review and meta-analysis. *Psychiatry Clin. Neurosci.* **76**, 292–302 (2022).

### Author contributions

M.E., K.M. and H.L. conducted the literature searches and analyzed the results. A.K., H.F.J., K.K., M.D.V., D.G. and C.M. oversaw the entire review process, including manuscript review. All authors reviewed the manuscript.

### Funding

Open access funding provided by Swiss Federal Institute of Technology Zurich.

### Competing interests

The authors declare no competing interests. M.E. serves as an Associate Editor for *npj Biosensing* and had no involvement in the review or editorial handling of this manuscript.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43856-025-01234-6>.

**Correspondence** and requests for materials should be addressed to Mohamed Elgendi or Carlo Menon.

**Peer review information** : *Communications Medicine* thanks Raymundo Cassani and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025