A scoping review on monitoring mental health using smart wearable devices

Nannan Long¹,², Yongxiang Lei³, Lianhua Peng²,⁴, Ping Xu⁵ and Ping Mao¹,⁶,*

¹ Department of Nursing, The Third Xiangya Hospital, Central South University, Changsha 410013, China
² Xiangya Nursing School, Central South University, Changsha 410031, China
³ Department of Mechanical Engineering, Politecnico di Milano, Milan 10056, Italy
⁴ Affiliated Hospital of Jinggangshan University, Jianggangshan 343100, China
⁵ ZiBo Hospital of Traditional Chinese and Western Medicine, Zibo 255020, China
⁶ Hunan Key Laboratory of Nursing, Changsha 410013, China

* Correspondence author: Email: ping.mao@csu.edu.cn.

Abstract: With the continuous development of the times, social competition is becoming increasingly fierce, people are facing enormous pressure and mental health problems have become common. Long-term and persistent mental health problems can lead to severe mental disorders and even death in individuals. The real-time and accurate prediction of individual mental health has become an effective method to prevent the occurrence of mental health disorders. In recent years, smart wearable devices have been widely used for monitoring mental health and have played an important role. This paper provides a comprehensive review of the application fields, application mechanisms, common signals, common techniques and results of smart wearable devices for the detection of mental health problems, aiming to achieve more efficient and accurate prediction for individual mental health, and to achieve early identification, early prevention and early intervention to provide a reference for improving the level of individual mental health.

Keywords: artificial intelligence; mental health; psychological prediction; signal indicator

Abbreviation: ACC: Accelerometer; ANN: Artificial neural network; BP: Blood pressure; BVP: Blood volume pulse; CNN: Convolutional neural network; DNN: Deep neural network; ECG:
Electrocardiograph; EDA: Electrodermal activity; EEG: Electroencephalography; EMG: Electromyogram; GSR: Galvanic skin response; HR: Heart rate; HRV: Heart rate variability; IBI: Interbeat interval; KNN: K-nearest neighbor; LASSO: Least absolute shrinkage selection operator; PCA: Principal component analysis; PPG: Photoplethysmography; RF: Random forest; RR: Respiration rate; RBF: Radial basis function kernel classifier; ST: Skin temperature; SVM: Support vector machine

1. Introduction

Mental health problems are also known as mental health disorders, and they have become an important factor affecting individual growth and social stability [1]. According to the “Global Burden of Disease Study”, the global prevalence of mental health disorders has always been above 10%, and about 450 million people have mental health disorders [2,3]. Long-term and persistent mental health problems increase the risk of cardiovascular disease, which is the leading cause of increased mortality worldwide [4]. Mental health disorders not only damage the physical and mental health of individuals, but they also increase the overall medical burden on society [5]. It is predicted that the total global medical expenditure for mental health problems will reach 16 trillion US dollars by 2030 [6]. Therefore, it is imperative to detect the mental health condition of individuals at an early stage and prevent the occurrence of mental health problems.

Smart sensors can accurately detect subtle changes (such as insomnia, headache, rapid heartbeat and muscle tension) in the body caused by stress in order to prevent and reduce the adverse effects of mental health problems [7,8]. With the increasing need for mental health care, the application of portable smart devices for monitoring mental health has received widespread attention and has been welcomed by most people [9]. At present, there are many studies on the use of wearable devices for mental health monitoring. Previous reviews have integrated the use of wearable sensors to detect stress levels, and there is a lack of exhaustive reviews on detection topics such as mental health problems or mental illness [8,10–13]. However, although some of the studies covered comprehensive information, the classification of detected signals were not described in detail in these reviews [14–17]. To fill this gap, this paper analyzes and integrates a large number of studies, covering the types of equipment used for data collection, common locations of sensors, common signals, advantages and disadvantages of technology, types of mental health states, etc. In this section, we discuss some existing studies. Abbreviations of nomenclatures used in this review paper are enumerated at the beginning of the article.

First, smart wearable devices have been widely used in the study of classifying and identifying individuals’ daily stress and mental fatigue. Betti et al. [18] measured the electroencephalography (EEG), electrocardiography (ECG) and electrodermal activity (EDA) of healthy adults using wearable devices and a support vector machine (SVM) classification algorithm for statistical analysis; the results show that the accuracy of the classification algorithm based on salient features can reach 86% for detecting individual stress levels. Goumopoulos et al. [19] proposed a method to detect mental fatigue; they used a wearable device to collect the heart rate variability (HRV) of healthy subjects, and then selected an appropriate machine learning model to predict mental fatigue with high accuracy. Additionally, the role of wearable devices in the prediction and classification of mental health assessment and emotional states cannot be ignored. Sano et al. [20] explored the relationships between the academic performance, sleep quality, stress perception and mental health of college students using wearable devices and smartphones; they found that college students’ academic performance, sleep quality and stress perception can significantly predict their mental health level, and that the accuracy
rate of the algorithm is as high as 92%. Costa et al. [21] found that wearable devices can acquire and analyze social signals and identify the current negative emotions of individuals, and that, when it identifies a negative emotion, the device activates the emotional regulation system to help the individual stabilize their emotions. Overall, wearable devices have been widely used in the field of mental health with strong practicability. The main contributions of this paper are given as follows:

1) Recent artificial intelligence (AI)-based wearable devices methods and techniques are reviewed in this paper. This study comprehensively explores the application of AI-based wearable devices in the field of mental health, focusing on stress levels, mental health problems and mental illness.

2) This study covers comprehensive and detailed information, including types and application mechanisms for AI-based wearable devices, common locations of sensors, common signals, the advantages and disadvantages of technology and types of mental health states.

3) The articles included in this paper are from 2017 to 2022, and the articles are relatively new and cutting-edge.

The remainder of this paper is organized as follows. Section 2 introduces the review of body signals and mental health. Section 3 summarizes the methodology and Section 4 gives the literature survey. Finally, application, discussion and conclusion are given in Sections 5, 6 and 7.

2. Body signals and mental health

The autonomic nervous system has two branches: the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). The SNS can rapidly stimulate some vital organs (such as the heart, lungs, and sweat glands), and the PNS can inhibit this excitation; they complement the regulation of important bodily functions and maintenance of physiological homeostasis [22]. When an individual is stressed, the individual will activate the protective mechanism; the SNS will be activated, the heart rate will increase, the breathing will increase, sweat will be secreted and other physiological systems will also be affected [23,24]. Additionally, the hypothalamus axis is a hormonal system that is initiated in response to stressors and drives a cascade of endocrine responses, culminating in the secretion of cortisol [22]. Therefore, individual health, especially mental health, can be known by observing the activity of the SNS and hypothalamus axis. Finally, language signals can also reflect an individual's mental health condition. Language is the basis for inferring individual thought processes; it provides a special perspective for thinking observations [25]. When an individual is in a state of tension, the respiratory rate increases and the glottis pressure increases, prompting an increase in the pitch of the vocalized part [25]. Therefore, clinicians mainly diagnose and treat mental illness through language and pronunciation [26]. As mentioned above, subtle changes in the body caused by stress can be detected by sensors connected to the body and hardware devices to accurately predict mental illness [27]. Therefore, AI-based wearable devices can collect one or more signals of the individual to identify and analyze the mental health level.

3. Methodology

3.1. Search strategy

This review incorporated searches for relevant papers in the databases of Web of Science, Embase,
MEDLINE and Engineering Village. The selected papers ranged from 2017 and 2022, which covers the recent developments in this area. The basic sets of keywords were “Psychological assessment”, “psychological prediction”, “psychological monitoring”, “mental health”, “psychological health”, “mental illness”, “psychological disorders” and “wearable device”. A total of 1537 papers were retrieved by using all keywords, and 25 papers were finally included. The search methodology for this article was designed to provide a detailed review of published research papers on wearable sensor-based mental health detection. Figure 1 shows a flowchart describing the search strategy used for this review article.

3.2. Inclusion criteria

1) The date of publication of the selected papers was between 2017 and 2022. 2) Conference papers and journal papers were included. 3) The research direction for the paper was mental health detection (focusing on stress, mental health problems, mental illness, etc.) and AI-based wearable devices for data collection.

3.3. Exclusion criteria

1) Research published in a language other than English was excluded. 2) Studies structured as editorials, press releases, research highlights or letters, review articles, reviews and technical papers were excluded. 3) Duplicate papers, related review papers and papers not related to the topic were excluded.

Figure 1. Flow diagram of the systematic search strategy.
4. Literature survey

4.1. Wearable devices for mental health monitoring

We use smart/embedded wearable devices, such as smartwatches, that collect signals from the body [28]. This data can be processed by sensors on smart/embedded devices. WiFi is used to send the collected data to cloud servers [7]. In cloud computing, advanced AI models can be applied to mental health monitoring big data to generate simple knowledge through the use of expensive processing and advanced AI algorithms [29], such as Welch's algorithm [30], SVM algorithms [31] and period graphs [32], for the analysis of data. Using these AI algorithms, then by analyzing the information, we will be able to extract sensor data with the appropriate classification characteristics and then classify their mental health states [33]. Artificial intelligence algorithms such as information and SVMs are used again. The main function of an analog classifier is to classify different classifiers [34]. By ensuring the rapidity of calculation, feature patterns are transformed into related classes and the classes most suitable for a single feature are predicted. During the classification process, each feature will be analyzed to find the most suitable feature [35]. They will be analyzed separately to predict stresses. Once the classification is complete, the smart device displays their mental health rating on the screen. Therefore, people can improve their mental health through self-management (as shown in Figure 2). The virtual reality-augmented reality (VR-AR) based digital twin is the critical element for the intelligent development of AI-based wearable devices [36]. Finally, all of the elements are integrated into a smart device’s ecology for the future direction framework.

![Intelligent ecology flowchart of AI-based wearable devices with data storage, transaction, interaction and communication networks.](image)

**Figure 2.** Intelligent ecology flowchart of AI-based wearable devices with data storage, transaction, interaction and communication networks.
Mental health detection systems based on wearable devices mainly consist of sensors, data acquisition, data pre-processing, feature extraction, machine learning algorithms and mental health condition classification. The general scheme of mental health prediction is shown in Figure 3. Section 4 will systematically introduce the current related research work according to the workflow of the wearable mental health detection system, as well as classify, compare and statistically analyze the existing research results.

4.2. Data acquisition devices

Currently, the commonly used signals in mental health detection are HRV, EEG, galvanic skin response (GSR), skin temperature (ST), blood pressure (BP), sleep, salivary cortisol and blood oxygen saturation (SpO2). In most studies, researchers use multiple indicators to accurately reflect an individual's mental health. At the same time, the wearable device can also collect indicators such as steps, activity category, posture and activity trajectory. According to the needs of users and the functions achieved by the device, the device will be worn on different parts of the body. Common wearing positions are shown in Figure 4.
4.3. Pre-processing and feature extraction

To solve the noise phenomenon in the data acquisition module, data pre-processing is required to reduce the impact of noise on the performance of mental health detection. Common filtering algorithms in mental health detection systems are wavelet transforms, the Kalman filter, etc. These filters can filter out high-frequency noise signals in the data [37,38].

The pre-processed data need to extract corresponding features for training the mental health detection model. Among them, the common feature extraction method is to use the sliding window algorithm to extract the features in the corresponding window, and then put the trained features into the trainer for training to form a mental health detection model. Table 1 summarizes the features extracted by the feature extraction module of an existing detection system. The commonly used feature values for feature extraction are Mean, Standard Deviation (StDev), Skew, Root Mean Square (RMS), Kurtosis, Quart1, Quart3, Inter-Quartile Range (IRQ), Sum, Frequency, etc.

<table>
<thead>
<tr>
<th>References</th>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[19,39–51]</td>
<td>Mean</td>
<td>( \text{Mean} = \sum_{i} s_{i}^{2} / n )</td>
</tr>
<tr>
<td>[19,39–47,49–51]</td>
<td>StDev</td>
<td>( sd = \sqrt{\frac{\sum_{i=1}^{N} (x_{i} - \mu)^{2}}{(N - 1)}} )</td>
</tr>
<tr>
<td>[40,45]</td>
<td>Skew</td>
<td>( G_{1} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left( \frac{x_{i} - \bar{x}}{s} \right)^{3} )</td>
</tr>
<tr>
<td>[19,40,43]</td>
<td>RMS</td>
<td>( \text{rmA}(s) = \sqrt{\sum s_{j}^{2} / x} )</td>
</tr>
<tr>
<td>[39,40,45,51]</td>
<td>Kurtosis</td>
<td>( K = \mu / \sigma )</td>
</tr>
<tr>
<td>[39,40,45]</td>
<td>Quart1</td>
<td>Median of lower 25% of the values</td>
</tr>
<tr>
<td>[39,40,45]</td>
<td>Quart3</td>
<td>Median of up 25% of the values</td>
</tr>
<tr>
<td>[39,40,45,51]</td>
<td>IRQ</td>
<td>Difference between Quart1 and Quart3</td>
</tr>
<tr>
<td>[41,42,44]</td>
<td>Sum</td>
<td>( S = \sum_{i=1}^{N} w_{i} )</td>
</tr>
<tr>
<td>[45]</td>
<td>Variance</td>
<td>( sd^{2} = 1 / n \sum_{i=1}^{n} (x_{i} - \mu)^{2} )</td>
</tr>
<tr>
<td>[43]</td>
<td>Coefficient of variation</td>
<td>( CV = sd / \mu )</td>
</tr>
<tr>
<td>[45]</td>
<td>Inverse coefficient of variation</td>
<td>( ICV = \mu / sd )</td>
</tr>
<tr>
<td>[18,19,41,43,46,47, 50–52]</td>
<td>Frequency</td>
<td>Peak frequency, low-frequency power (LF), high-frequency power (HF), very-low-frequency power (VLF), prevalent low frequency (pLF), prevalent high frequency (pHF), the ratio of LF to HF (LF/HF), (From Lomb–Scargle) LF, HF and LF/HF</td>
</tr>
</tbody>
</table>
Table 2. Comparison analysis summary of several existing machine learning techniques.

<table>
<thead>
<tr>
<th>Machine learning algorithms</th>
<th>Definition</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Machine (SVM)</td>
<td>In order to better find the hyperplane that can divide the feature space, the low-dimensional data are mapped to the high-dimensional plane</td>
<td>Works well with any complex problem, can scale high-dimensional data, risk of overfitting is lower; It shows high robustness and strong generation ability with low overfitting</td>
<td>Sensitive to missing data; It is difficult to implement large-scale training samples; There are difficulties in solving the multi-classification problem</td>
</tr>
<tr>
<td>K-Nearest Neighbor (KNN)</td>
<td>In the feature space, if k samples near a sample belong to a category, that sample also belongs to that category</td>
<td>Simple thought, mature theory; Can be used for multiple classification, high accuracy, no assumptions on the data, not sensitive to outliers</td>
<td>Poor generalization ability and slow prediction speed</td>
</tr>
<tr>
<td>Random Forests (RF)</td>
<td>A classifier for multiple decision trees</td>
<td>It can handle high-dimension data, has strong model generalization ability, and is fast in training. It can balance errors in unbalanced data sets, maintain accuracy if a large part of features is lost</td>
<td>In noisy classification or regression problems, overfitting and attribute weights produced based on data of different attribute values are not credible</td>
</tr>
<tr>
<td>Artificial Neural Network (ANN)</td>
<td>A mathematical model for distributed parallel information processing that mimics the features of human neural network</td>
<td>It has a high accuracy of classification, strong parallel distribution processing ability, strong distribution storage and learning ability, adaptability, strong robustness and fault tolerance to noise nerve</td>
<td>Many parameters are required, which are not suitable for small samples, high calculation cost, long learning process, and intermediate results cannot be observed</td>
</tr>
<tr>
<td>Logical Regression (LR)</td>
<td>The result of linear regression is mapped between 0 and 1 by a sigmoid function. The result of the mapping is the probability that the data sample point belongs to a certain category</td>
<td>The computational cost is low, the implementation is easy to understand, the robustness to small noise in the data is good, and it is not particularly affected by slight multicollinearity</td>
<td>It is easy to underfit, and the performance effect is not good when the classification accuracy is not high, the data features are missing or the feature space is large</td>
</tr>
<tr>
<td>Decision Tree (DT)</td>
<td>A prediction model is a mapping between object attributes and object values</td>
<td>Easy to understand and explain, visualization of tree structure, less data required for training, fast analysis speed, and strong ability to process samples with missing attributes</td>
<td>Overfitting, poor generalization ability and unstable calculation results</td>
</tr>
</tbody>
</table>

Continued on next page
### Machine learning algorithms

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Definition</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian Network</td>
<td>Graphical networks based on probabilistic inference</td>
<td>The process is simple and fast, efficient for multi-classification problems, and requires a smaller sample size</td>
<td>Sample attributes do not work well when they are associated and are sensitive to input forms</td>
</tr>
<tr>
<td>Linear Discriminant Analysis (LDA)</td>
<td>The labeled data is projected to a lower dimensional space so that the points after projection are distinguished by category. Points of the same category are closer to each other in the space after projection, and the differences are as far as possible</td>
<td>It can reflect the differences between samples and use the prior knowledge experience of categories in dimension reduction</td>
<td>Easy to fit, restricted by sample type</td>
</tr>
<tr>
<td>Principal Component Analysis (PCA)</td>
<td>A method of multivariate statistical analysis in which several variables are transformed linearly to select a smaller number of important variables</td>
<td>Eliminate the mutual influence between indicators, small workload</td>
<td>The amount of information must be kept at a high level after dimensionality reduction, and the interpretation of principal components is usually fuzzy</td>
</tr>
</tbody>
</table>

### 4.4. Machine learning techniques

Machine learning technology refers to the method of using some special formulas to guide the computer to use the known data to build a model that can make predictions for new situations. Common machine learning algorithms for mental health detection include SVMs [18,19,39,43,44,48,51–55] k-nearest neighbor (KNN) algorithms [19,48,51,53], random forests (RF) [40,45,51], artificial neural networks (ANNs) [39], logistic regression (LR) [19,42,56,50,51], decision trees (DTs), Bayesian networks (BNs) [48], linear discriminant analysis (LDA) [51] and principal component analysis (PCA) [47,51]. The SVM is the most used algorithm for mental health detection among them. This paper summarizes the artificial intelligence algorithms involved in mental health detection and briefly describes their definitions, advantages and disadvantages, as shown in Table 2.

### 5. Application

Mental health problems are often accompanied by obvious body responses, and wearable sensors play an important role in mental health and its related applications. They can collect an individual's EEG, heart rate (HR), HRV, GSR, BP, body temperature, respiratory rate, etc., to monitor mental health. This section summarizes and discusses the signals commonly used in mental health detection, and briefly summarizes the relevant research on AI-based wearable devices in mental health detection over the past five years. The main contents include mental health disease types, parameters, results, etc., as shown in Table 3.
Table 3. Mental health applied research model and summary of results.

<table>
<thead>
<tr>
<th>Study</th>
<th>Type</th>
<th>Device</th>
<th>Signals</th>
<th>Model</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can et al., 2019 [51]</td>
<td>Stress</td>
<td>Samsung Gear S, S2 and Empatica E4 wristband</td>
<td>End Diastolic Velocity, HRV and ACC HRV, EDA, ECG</td>
<td>PCA, LDA, SVM, KNN, RF and LR</td>
<td>The accuracy of the multi-layer perceptron algorithm is 92.19%.</td>
</tr>
<tr>
<td>Betti et al., 2018 [18]</td>
<td>Stress</td>
<td>Mobile EEG headset, EEG and Chest belt</td>
<td>Salivary cortisol</td>
<td>SVM</td>
<td>The classification algorithm based on salient features has an accuracy of 86%.</td>
</tr>
<tr>
<td>Egilmez et al., 2017 [40]</td>
<td>Stress</td>
<td>LG smartwatch</td>
<td>GSR</td>
<td>RF</td>
<td>The pressure detection F-value of the system is 88.8%.</td>
</tr>
<tr>
<td>Gjoreski et al., 2019 [39]</td>
<td>Stress</td>
<td>Empatica device’s wristband</td>
<td>HR, BVP, IBI, ST and ACC</td>
<td>SVM</td>
<td>Experiments with 55 days of real data showed that the method detected stress events with 95% accuracy.</td>
</tr>
<tr>
<td>Ahn et al., 2019 [52]</td>
<td>Stress</td>
<td>Head-mounted electrode</td>
<td>ECG, EEG</td>
<td>SVM</td>
<td>The accuracy of the classification of pressure states using SVM technology and the 5-fold cross validation of EEG and HRV features was 87.5%. Mental stress index is sensitive to acute stress and can predict the level of association between normal individuals and the stress group with about 97% accuracy.</td>
</tr>
<tr>
<td>Wu et al., 2019 [50]</td>
<td>Stress</td>
<td>Smart shirt</td>
<td>Cortisol, HRV</td>
<td>LR</td>
<td>Multi-sensor data fusion can better monitor individual stress levels.</td>
</tr>
<tr>
<td>Jesmin et al., 2020 [53]</td>
<td>Stress</td>
<td>Smartwatch</td>
<td>HRV, GSR and SpO2</td>
<td>SVM, KNN</td>
<td>Using machine learning technology, the accuracy of Class III and Class II classification problems reached 81.65 and 93.20%, respectively, and using deep learning technology, the accuracy of Class III and Class II classification problems reached 84.32 and 95.21%, respectively. Several parameters of the HRV under pressure were significantly different from baseline.</td>
</tr>
<tr>
<td>Bobade et al., 2020 [41]</td>
<td>Stress</td>
<td>Chest-worn device, RespiBAN Professional and a wrist-worn device Empatica E4</td>
<td>ACC, ECG, BVP, ST, RR, EMG and EDA</td>
<td>Machine Learning</td>
<td></td>
</tr>
<tr>
<td>Silva et al., 2020 [49]</td>
<td>Stress</td>
<td>Smartwatch</td>
<td>HRV</td>
<td>ANN</td>
<td></td>
</tr>
</tbody>
</table>

*Continued on next page*
<table>
<thead>
<tr>
<th>Study</th>
<th>Type</th>
<th>Device</th>
<th>Signals</th>
<th>Model</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al., 2020 [42]</td>
<td>Stress</td>
<td>Empatica E4 wristband</td>
<td>GSR</td>
<td>LR</td>
<td>The overall classification accuracy of the model was 85.3%, and the cross-validation accuracy was 83.2%. The accuracy of 10-fold cross-validation was 94.55%, while the accuracy of subject cross-validation was 85.71%. In daily settings, the system had 81.82% accuracy in assessing stress. HR increased with both sensory (olfactory) and cognitive stimulation, and there was a similar decrease in HRV. Overall, HRV characteristics calculated from wearable vest data (with higher temporal resolution) proved to be more effective in predicting stress and anxiety. Method had an average accuracy of 92 and 89% in distinguishing the stress levels of the experimental group and the control group, respectively. The sensors were 78.3% accurate in classifying students as high or low stress groups, and 87% accurate in classifying students as high or low mental health groups. There was significant variability in skin resistance and HRV under relaxation and stimulation. The use of wearable long-term speech-to-social data sets to categorize various sound scenes in unconstrained and natural environments yielded good results. The classifier achieved 81.82% accuracy in detecting a high anxiety state.</td>
</tr>
<tr>
<td>Han et al., 2020 [48]</td>
<td>Stress</td>
<td>Smart wrist band, Empatica E4 wristband and a Shimmer ECG device</td>
<td>ECG, PPG and GSR</td>
<td>KNN, SVM, and BN</td>
<td>The accuracy of 10-fold cross-validation was 94.55%, while the accuracy of subject cross-validation was 85.71%. In daily settings, the system had 81.82% accuracy in assessing stress.</td>
</tr>
<tr>
<td>Tonacci et al., 2019 [62]</td>
<td>Stress</td>
<td>Shimmer™, Shimmer 3™</td>
<td>GSR, ECG</td>
<td>Pan Tompkins algorithm</td>
<td>(olfactory) and cognitive stimulation, and there was a similar decrease in HRV. Overall, HRV characteristics calculated from wearable vest data (with higher temporal resolution) proved to be more effective in predicting stress and anxiety. Method had an average accuracy of 92 and 89% in distinguishing the stress levels of the experimental group and the control group, respectively. The sensors were 78.3% accurate in classifying students as high or low stress groups, and 87% accurate in classifying students as high or low mental health groups. There was significant variability in skin resistance and HRV under relaxation and stimulation. The use of wearable long-term speech-to-social data sets to categorize various sound scenes in unconstrained and natural environments yielded good results. The classifier achieved 81.82% accuracy in detecting a high anxiety state.</td>
</tr>
<tr>
<td>Tiwari et al., 2019 [43]</td>
<td>Stress and anxiety</td>
<td>OMsignal smart-shirt and a Fitbit Charge 2 HRV smart-bracelet</td>
<td>HRV</td>
<td>SVM</td>
<td>(olfactory) and cognitive stimulation, and there was a similar decrease in HRV. Overall, HRV characteristics calculated from wearable vest data (with higher temporal resolution) proved to be more effective in predicting stress and anxiety. Method had an average accuracy of 92 and 89% in distinguishing the stress levels of the experimental group and the control group, respectively. The sensors were 78.3% accurate in classifying students as high or low stress groups, and 87% accurate in classifying students as high or low mental health groups. There was significant variability in skin resistance and HRV under relaxation and stimulation. The use of wearable long-term speech-to-social data sets to categorize various sound scenes in unconstrained and natural environments yielded good results. The classifier achieved 81.82% accuracy in detecting a high anxiety state.</td>
</tr>
<tr>
<td>Patlar Akbulut et al., 2020 [47]</td>
<td>Stress and anxiety</td>
<td>Wrist-worn wearable device</td>
<td>ECG, GSR, ST, SpO2, Blood sugar and BP</td>
<td>PCA, Feed-forward neural network</td>
<td>(olfactory) and cognitive stimulation, and there was a similar decrease in HRV. Overall, HRV characteristics calculated from wearable vest data (with higher temporal resolution) proved to be more effective in predicting stress and anxiety. Method had an average accuracy of 92 and 89% in distinguishing the stress levels of the experimental group and the control group, respectively. The sensors were 78.3% accurate in classifying students as high or low stress groups, and 87% accurate in classifying students as high or low mental health groups. There was significant variability in skin resistance and HRV under relaxation and stimulation. The use of wearable long-term speech-to-social data sets to categorize various sound scenes in unconstrained and natural environments yielded good results. The classifier achieved 81.82% accuracy in detecting a high anxiety state.</td>
</tr>
<tr>
<td>Sano et al., 2018 [54]</td>
<td>Stress and mental health condition</td>
<td>Q-sensor and Motion logger</td>
<td>SC, ST, ACC and Motion logger</td>
<td>SVM, LASSO and RBF</td>
<td>(olfactory) and cognitive stimulation, and there was a similar decrease in HRV. Overall, HRV characteristics calculated from wearable vest data (with higher temporal resolution) proved to be more effective in predicting stress and anxiety. Method had an average accuracy of 92 and 89% in distinguishing the stress levels of the experimental group and the control group, respectively. The sensors were 78.3% accurate in classifying students as high or low stress groups, and 87% accurate in classifying students as high or low mental health groups. There was significant variability in skin resistance and HRV under relaxation and stimulation. The use of wearable long-term speech-to-social data sets to categorize various sound scenes in unconstrained and natural environments yielded good results. The classifier achieved 81.82% accuracy in detecting a high anxiety state.</td>
</tr>
<tr>
<td>Fraiwan et al., 2018 [65]</td>
<td>Mental health condition</td>
<td>Smart glove</td>
<td>ST and HRV HRV algorithm</td>
<td>HRV algorithm</td>
<td>(olfactory) and cognitive stimulation, and there was a similar decrease in HRV. Overall, HRV characteristics calculated from wearable vest data (with higher temporal resolution) proved to be more effective in predicting stress and anxiety. Method had an average accuracy of 92 and 89% in distinguishing the stress levels of the experimental group and the control group, respectively. The sensors were 78.3% accurate in classifying students as high or low stress groups, and 87% accurate in classifying students as high or low mental health groups. There was significant variability in skin resistance and HRV under relaxation and stimulation. The use of wearable long-term speech-to-social data sets to categorize various sound scenes in unconstrained and natural environments yielded good results. The classifier achieved 81.82% accuracy in detecting a high anxiety state.</td>
</tr>
<tr>
<td>Chen et al., 2018 [68]</td>
<td>Mental health condition</td>
<td>Smartwatch</td>
<td>Speech</td>
<td>Transfer learning</td>
<td>(olfactory) and cognitive stimulation, and there was a similar decrease in HRV. Overall, HRV characteristics calculated from wearable vest data (with higher temporal resolution) proved to be more effective in predicting stress and anxiety. Method had an average accuracy of 92 and 89% in distinguishing the stress levels of the experimental group and the control group, respectively. The sensors were 78.3% accurate in classifying students as high or low stress groups, and 87% accurate in classifying students as high or low mental health groups. There was significant variability in skin resistance and HRV under relaxation and stimulation. The use of wearable long-term speech-to-social data sets to categorize various sound scenes in unconstrained and natural environments yielded good results. The classifier achieved 81.82% accuracy in detecting a high anxiety state.</td>
</tr>
<tr>
<td>Wen et al., 2020 [44]</td>
<td>Anxiety</td>
<td>The MP150 multi-channel physiological recorder</td>
<td>HR (ECG electrodes)</td>
<td>SVM</td>
<td>(olfactory) and cognitive stimulation, and there was a similar decrease in HRV. Overall, HRV characteristics calculated from wearable vest data (with higher temporal resolution) proved to be more effective in predicting stress and anxiety. Method had an average accuracy of 92 and 89% in distinguishing the stress levels of the experimental group and the control group, respectively. The sensors were 78.3% accurate in classifying students as high or low stress groups, and 87% accurate in classifying students as high or low mental health groups. There was significant variability in skin resistance and HRV under relaxation and stimulation. The use of wearable long-term speech-to-social data sets to categorize various sound scenes in unconstrained and natural environments yielded good results. The classifier achieved 81.82% accuracy in detecting a high anxiety state.</td>
</tr>
<tr>
<td>Study</td>
<td>Type</td>
<td>Device</td>
<td>Signals</td>
<td>Model</td>
<td>Result</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------</td>
<td>---------------------------------------</td>
<td>--------------------------</td>
<td>---------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>O’Brien et al., 2017 [56]</td>
<td>Depression</td>
<td>Wrist-mounted device</td>
<td>ACC</td>
<td>LR</td>
<td>General physical activity was significantly reduced in depressed older adults compared with healthy controls. Participants' ecological transient assessment self-report and passive perception (sensor data) were highly correlated between physical activity, mood and sleep levels; Depression detector verified the feasibility of grouping and classification with an accuracy rate of 96.00% (StDev = 2.76). Patients with depression were accurately detected in 86.7% of the cases, while those without depression were found in 91.9% of the cases.</td>
</tr>
<tr>
<td>Narziev et al., 2020 [67]</td>
<td>Depression</td>
<td>Smartwatch</td>
<td>Motion logger and HR</td>
<td>Machine learning</td>
<td>Participants' ecological transient self-report and passive perception (sensor data) were highly correlated between physical activity, mood and sleep levels; Depression detector verified the feasibility of grouping and classification with an accuracy rate of 96.00% (StDev = 2.76). Patients with depression were accurately detected in 86.7% of the cases, while those without depression were found in 91.9% of the cases.</td>
</tr>
<tr>
<td>Zanella-Calzada et al., 2019 [45]</td>
<td>Depression</td>
<td>Activwatch</td>
<td>Motion Logger</td>
<td>RF</td>
<td>Participants' ecological transient self-report and passive perception (sensor data) were highly correlated between physical activity, mood and sleep levels; Depression detector verified the feasibility of grouping and classification with an accuracy rate of 96.00% (StDev = 2.76). Patients with depression were accurately detected in 86.7% of the cases, while those without depression were found in 91.9% of the cases.</td>
</tr>
<tr>
<td>Gentili et al., 2017 [55]</td>
<td>Bipolar disorder</td>
<td>Sensorized t-shirt</td>
<td>HRV, respiration and ACC</td>
<td>SVM</td>
<td>Participants' ecological transient self-report and passive perception (sensor data) were highly correlated between physical activity, mood and sleep levels; Depression detector verified the feasibility of grouping and classification with an accuracy rate of 96.00% (StDev = 2.76). Patients with depression were accurately detected in 86.7% of the cases, while those without depression were found in 91.9% of the cases.</td>
</tr>
<tr>
<td>Zheng et al., 2019 [46]</td>
<td>Emotion recognition</td>
<td>Head/wearable eye-tracking glasses, EmotionMeter hardware</td>
<td>EEG and Eye-movement signal</td>
<td>Multi-mode deep neural network</td>
<td>Participants' ecological transient self-report and passive perception (sensor data) were highly correlated between physical activity, mood and sleep levels; Depression detector verified the feasibility of grouping and classification with an accuracy rate of 96.00% (StDev = 2.76). Patients with depression were accurately detected in 86.7% of the cases, while those without depression were found in 91.9% of the cases.</td>
</tr>
<tr>
<td>Goumopoulos et al., 2022 [19]</td>
<td>Mental fatigue</td>
<td>Zephyr HXM-BT, Chest-strap built-in ECG sensor</td>
<td>HRV</td>
<td>SVM, KNN and LR</td>
<td>Mental fatigue caused by cognitive load can be detected by wearable commercial devices and a single biomarker. Through the fatigue evaluation experiment of 10 healthy subjects, it was verified that the device can reliably detect the mental fatigue state by monitoring and analyzing EEG, ECG and proximal ST data.</td>
</tr>
<tr>
<td>Qiu et al., 2020 [38]</td>
<td>Mental fatigue</td>
<td>Head-mounted devices</td>
<td>ECG, EEG and ST</td>
<td>Wavelet transform, Kalman filtering</td>
<td>Mental fatigue caused by cognitive load can be detected by wearable commercial devices and a single biomarker. Through the fatigue evaluation experiment of 10 healthy subjects, it was verified that the device can reliably detect the mental fatigue state by monitoring and analyzing EEG, ECG and proximal ST data.</td>
</tr>
</tbody>
</table>

5.1. Mental health detection using EEG

EEG is a detection method that records the electrical activity of the cerebral cortex by connecting multiple electrodes to the scalp [57] to reflect the emotional changes of individuals to specific stimuli.
The stress state of the human body is mainly identified by the changes in the alpha and delta bands of the EEG. When people are under high levels of stress, alpha decreases and delta increases in the EEG, whereas, when they are relaxed or inactive, alpha increases and delta decreases [58]. Ahn et al. [52] used a binaural hanging device that combined EEG and ECG signals and found that the accuracy of EEG and ECG signals in continuous stress assessment was 87.5%. Additionally, based on the multi-modal emotion recognition framework of EmotionMeter, some scholars found that the individual's happiness, sadness, fear and neutral emotions can be distinguished and identified by EEG signals [46]. Finally, EEG combined with ECG signals, body temperature and other data can reliably detect the mental fatigue state of an individual, which provides support for the popularization and specific implementation of mental health assessments [38].

Table 3 shows the studies and their details on the use of EEG for mental health detection. EEG is one of the most effective physiological signals for detecting psychological stress, and the combination of ECG signals and machine learning applications can improve the accuracy of psychological stress detection.

5.2. Mental health detection using ECG

Stressors induce changes in myocardial activity, promote blood distribution to vital organs and cause changes in the HR and HRV [59]. ECG is a reliable technique that uses electrodes attached to the body to record the electrical potential produced by heart muscle cells to observe changes in HRV [59]. Studies have shown that HR and HRV signals are closely related to mental health problems (such as anxiety and depression) [60], and they have an important role in mental health detection. Tiwari et al. [43] found that HRV data obtained through smart t-shirts and smart bracelets can more accurately assess the degree of anxiety and depression of individuals. Second, HRV characteristics are closely associated with an individual’s clinical course and mood changes. Gentili et al. [55] showed that the HRV characteristics (including time domain, frequency domain, nonlinearity) of patients with bipolar disorder can significantly predict their clinical course and emotional changes. Additionally, Akbulut et al. [47] believe that the stress state of an individual can be accurately identified by signals such as the HRV obtained by wearable devices. Table 3 presents an overview and the details of mental health testing using ECG.

5.3. Mental health detection using GSR

Sweat is a biological fluid produced by sweat glands. Sweat glands are distributed almost all over the body, with high-density distribution in the palms, soles and forehead; they are only stimulated by the sympathetic nerves, which are activated under tension [14], so the level of sweat secretion is the body’s response to stress. Since sweat contains ions, small molecules and even macromolecular substances such as proteins, it has electrical conductivity [61]. Therefore, the EDA signal is obtained by connecting two electrodes on the skin surface, and the voltage between them is measured to monitor the stress. Sriamprakash et al. [33] collected the GSR, HRV and other indicators, extracted individual stress characteristics by using Welch's algorithm and periodograms, etc., and used SVM and KNN algorithms to classify levels of mental health. Tonacci et al. [62] captured the ECG and skin conductance (SC) signals of subjects by using wearable sensors and compared the effects of cognitive stress and olfactory stimuli on human mental health. Additionally, EDA and GSR also change with different emotions; this can effectively reflect the state of emotional calm or emotional arousal [63].
Evidence shows that the emotional state of patients with metabolic syndrome can be analyzed by using indicators such as the GSR, HRV and body temperature detected by wearable devices [47]. The above studies have confirmed that combining GSR with HRV, body temperature and other parameters can improve the prediction rate for individual stress, and accurately provide and detect information about individual emotional states.

5.4. Mental health detection using ST

High levels of stress can cause muscle tension, increase the HR and increase the work of the vital organs; additionally, the blood will be shunted from the limbs to directly act on the vital organs. In just a few minutes, the body surface temperature can change [64]. Most of the studies involved the collection of body temperature data to detect mental health conditions. Fraiwan et al. [65] designed a mobile mental health monitoring system for continuous monitoring of changes in ST, skin electrical impedance and HRV; they found that, when the body was exposed to relaxation or stimulation, the skin electrical resistance value and HRV exhibited significant variability. Engelnietherammer et al. [66] monitored the EDA and ST in subjects wearing a smart wristband (Bodymonitor); the results showed that being in a crowded environment triggers aversive emotional responses (i.e., stress, anger and fear) in subjects. Sano et al. [54] had 201 college students wear two sensors on each wrist: a Q-sensor (for measuring SC, ST, triaxial acceleration (ACC)) and motion recorder (for measuring acceleration and ambient light data); the study showed that wearable sensor eigenvalues are more accurate than cell phones for classifying poor mental health (87%) and stress (78.3%) when using machine learning techniques. In conclusion, the ST, combined with parameters such as EDA, is more accurate for detecting emotional changes and stress levels.

5.5. Mental health detection using speech and behavior

With the development of artificial intelligence, the natural language processing program of the computer can more accurately identify the subtle barriers in language before the onset of mental illness, achieving the effect of accurately predicting the mental illness [27]. Corcoran et al. [27] confirmed that the automatic learning of speech classifiers with reduced semantic consistency, increased semantic consistency differences and reduced use of case pronouns can predict psychotic episodes with an accuracy rate of 83%. Language signals help to improve the accuracy of mental disease prediction and have become a powerful tool for psychiatric diagnosis and treatment.

In addition, mental illness is associated with changes in an individual’s behavior, especially physical movement and social habits. AI-based wearable devices (accelerometers, global position system sensors) can detect these abnormal behaviors and effectively reduce the incidence of individual mental illness. Zanella-Calzada et al. [45] obtained data on motor activity recorded by wearable devices from a depression database, extracted the statistical features and then fed them to an RF classifier to detect subjects with depression; the results showed that those subjects with the presence of depression had an 86.7% likelihood of being correctly classified, while the specificity showed that those subjects with the absence of depression had a 91.9% likelihood of being classified with a correct response. Additionally, Narziev et al. [67] recruited 20 participants from four different levels of depression groups and established a machine learning model to automatically classify the categories of depression in a short period of time; the results of this pilot study revealed high correlations between participants’ ecological
momentary assessment self-reports and passive sensing (sensor data) in terms of physical activity, mood and sleep levels. All of the above studies have shown that the role of speech signals and behavioral signals in the detection of mental health states such as depression, mood disorders or schizophrenia cannot be ignored.

6. Discussion

Multiple studies have integrated the use of wearable sensors to detect stress levels, and there is a lack of exhaustive reviews on topics such as mental health or illness detection. However, some reviews comprehensively cover information, but do not provide details such as machine learning methods, types of mental health states and the classification of detection signals. To fill this gap, this paper presents an extensive review of mental health detection using wearable sensors.

The significant observations from this review are listed below.
- Most wearable sensors are placed on the wrist (such as smartwatches and smart wristbands), chest (smart t-shirts, chest straps, etc.) and head (head-mounted devices, etc.) for data collection.
- The Empatica, EmotionMeter and SHIMMER platforms have been popularly used for data collection.
- HRV, EEG, GSR and ST are the most distinctive signals for detecting mental health conditions.
- SVM, LR and KNN have been widely used in the classification of mental health conditions.
- Most of the disease types studied focused on psychological stress levels and mental health problems (anxiety and depression).

The identified challenges are described below.
- Compliance: The collection of subjects' physiological indicators and daily activity data is affected by individual living habits, which makes the detection process more complicated and thus degrades the performance of the mental health detection system. Therefore, the mental health assessment based on wearable devices can only do a rough screening and cannot directly assess the disease accurately.
- Singleness of indicators: most studies involve the collection of only a single category of signals, such as physiology, speech and behavior, while fewer collect a combination. Since different categories of data and signals reflect different aspects of mental health, the integrated collection of all types of signals is essential for the accurate screening of mental health problems.

7. Conclusions

Globally, mental health problems are still serious social public health problems, and mental health has always been detected via subjective and time-consuming methods. Wearable sensors have great potential as a continuous monitoring system for mental health. As the needs for mental health care services increase, so does the production of AI-based wearable devices. This paper reviewed the types of equipment, common locations of sensors, common signals, advantages and disadvantages of technology and types of mental health states in previous studies. First of all, the most common types of states in mental health testing are psychological stress and mental health problems (anxiety and depression), and most wearable sensors are placed on the wrist, chest or head. Additionally, it was found that there are many AI-based wearable devices on the market that mainly collect physiological signals for mental health detection, and that there are few studies on detecting mental health with
speech signals and behavioral signals. Among the physiological signals, the HRV, EEG, ST and GSR signals are the most effective parameter indicators for detecting the mental health condition and this conclusion is helpful for researchers to create devices to improve people's mental health level. Finally, this paper integrated important information, such as the technical advantages and limitations of previous studies; it was found that the most common classifiers are SVM, KNN and LR algorithms, and that these algorithms provide the most reliable results for mental health detection. In the future, with the continuous innovation of science and technology, deterministic annealing neural networks [69,70], convolutional neural networks [71–73], the logarithmic descent direction algorithm [74], Laplacian-Hessian regularization semi-supervised extreme learning machine [75,76] and back-propagation neural networks [77] are also expected to be applied and explored in the field of mental health detection. In a word, the goal of mental health testing is to develop an efficient and affordable model with high accuracy. This review will help future researchers design optimal wearable devices to achieve their goals for mental health detection.

Acknowledgement

This work is supported by the Changsha Natural Science Foundation No. kq2202422 and the Chinese Scholarship Council (CSC) under grant No. 202006370101.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References


