applied sciences

MDPI

Systematic Review

# Capturing Mental Workload Through Physiological Sensors in Human–Robot Collaboration: A Systematic Literature Review

**Eduarda Pereira [1,*], Luis Sigcha [2], Emanuel Silva [1], Adriana Sampaio [3], Nuno Costa [4] and Nélson Costa [1]**

1 ALGORITMI Research Center/LASI, University of Minho, 4800-058 Guimarães, Portugal; id10249@alunos.uminho.pt (E.S.); ncosta@dps.uminho.pt (N.C.)
2 Department of Physical Education and Sports Science (PESS), Data-Driven Computer Engineering Research Centre (D2iCE) & Health Research Institute (HRI), University of Limerick, V94 T9PX Limerick, Ireland; luis.sigcha@ul.ie
3 Psychological Neuroscience Laboratory, CIPsi, School of Psychology, University of Minho, 4704-553 Braga, Portugal; adriana.sampaio@psi.uminho.pt
4 DTX-Colab, 4800-056 Guimarães, Portugal; nuno.costa@dtx-colab.pt
* Correspondence: id8976@alunos.uminho.pt

**Featured Application: The results of this study on mental workload have potential applications in developing adaptive robots that enhance safety and efficiency in human–robot collaboration within industrial settings.**

**Abstract:** Human–robot collaboration (HRC) is increasingly prevalent across various industries, promising to boost productivity, efficiency, and safety. As robotics technology advances and takes on more complex tasks traditionally performed by humans, the nature of work and the demands on workers are evolving. This shift emphasizes the need to critically integrate human factors into these interactions, as the effectiveness and safety of these systems are highly dependent on how workers cooperate with and understand robots. A significant challenge in this domain is the lack of a consensus on the most efficient way to operationalize and assess mental workload, which is crucial for optimizing HRC. In this systematic literature review, we analyze the different psychophysiological measures that can reliably capture and differentiate varying degrees of mental workload in different HRC settings. The findings highlight the crucial need for standardized methodologies in workload assessment to enhance HRC models. Ultimately, this work aims to guide both theorists and practitioners in creating more sophisticated, safe, and efficient HRC frameworks by providing a comprehensive overview of the existing literature and pointing out areas for further study.

**Keywords:** human factors; bio-signals; human–robot interaction; Industry 5.0

check for updates

Received: 13 January 2025
Revised: 10 March 2025
Accepted: 13 March 2025
Published: 18 March 2025

**Citation:** Pereira, E.; Sigcha, L.; Silva, E.; Sampaio, A.; Costa, N.; Costa, N. Capturing Mental Workload Through Physiological Sensors in Human–Robot Collaboration: A Systematic Literature Review. *Appl. Sci.* **2025**, *15*, 3317. https://doi.org/10.3390/app15063317

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

The advent of robotic technologies has catalyzed a paradigm shift in industrial operations worldwide. With their deployment, tasks once deemed dangerous or laborious for humans are now executed with robotic precision, freeing human operators to engage in more cognitively demanding activities that necessitate creative and analytical thinking. This symbiotic relationship between humans and robots has given rise to new production processes that prioritize human–robot interaction (HRI), recognizing its impact on both efficiency and safety in automated systems [1–3].

The integration of human factors in HRI is of paramount importance for understanding the cognitive and physical interplay between humans and robots. As task complexity

increases and collaborative robotic environments become more diverse, a comprehensive approach is required to ensure both efficiency and operator well-being. While significant progress has been made in regulating the physical safety of human–robot collaboration (HRC), the cognitive and emotional dimensions of these interactions remain underexplored. Addressing these aspects is crucial, particularly in dynamic industrial settings, where operators are exposed to mental workload, fatigue, and stress, which can significantly impact their performance and overall safety [4,5]. To regulate physical safety in HRI, international standards such as ISO 10218-2 [6] and ISO/TS 15066 [7] have been developed. These standards establish limitations on force, speed, and proximity, ensuring that robotic systems operate within safe parameters when interacting with human workers. Additionally, ISO/TS 15066 introduces biomechanical limits to mitigate the risk of injury, defining acceptable impact forces for direct human–robot contact. While these regulations represent a significant advancement in physical safety, they do not address cognitive ergonomics' risks, leaving a gap in the understanding of mental workload and its impact on HRC [4,5,8].

Mental workload is the cognitive effort required by an individual to achieve a certain level of task performance. This multifaceted construct is influenced by a variety of factors, including task complexity, robot capabilities, and environmental conditions [9]. However, the human factor itself is highly intricate, involving numerous interdependent variables, which makes it challenging to establish a universal and standardized definition of mental workload [10].

On this topic, three states of mental workload can be considered: mental overload, mental underload, and optimal mental workload [11]. Ideally, all employees would have an optimal mental workload level, that is, an appropriate amount and type of work. However, in multitasking environments such as industry scenarios, the mental workload can approach or exceed the upper limit of the worker's mental capacity, and it is known that high levels of mental workload can lead to reduced attention, decision-making errors, and slower response times, ultimately affecting both operator well-being and system efficiency [12]. On the opposite side of the spectrum is mental underload, characterized by a reduced or very easy amount of work [13], which is often associated with monotony and disengagement. Thus, assessing and managing the mental workload of workers is closely related to the ability to achieve multiple optimal performance goals.

An accurate measurement and management of mental workload is crucial for maintaining optimal performance and preventing accidents in HRC operations. Nevertheless, it is imperative to highlight the importance of assessing this, as consistent monitoring can identify critical thresholds beyond which operator efficiency declines and the risk of errors increases, thus ensuring both human and robotic components of the system can interact safely and effectively [14].

The ISO standard 10075-3 categorizes mental workload assessment methods into four distinct groups: physiology, subjective, performance, and job and task analysis [14]. Each category utilizes a unique approach to evaluate mental workload. Physiological measures, such as heart rate variability (HRV) and electrodermal activity (EDA), provide objective data on the operator's arousal and stress levels that can be used to infer mental workload. On the other hand, subjective measurements, such as the National Aeronautics and Space Administration Task Load Index (NASA-TLX), offer self-reported assessments of workload across various dimensions, including mental, physical, and temporal demands (as proposed by [15]). Performance methods, in turn, enable the evaluation of human mental and psychomotor performance under specific work conditions, helping to assess any performance decrements or variations due to mental workload effects. Lastly, job and task analysis methods examine task elements, physical and psychosocial work conditions, environmental factors, and the organization of the work process as potential sources

of mental workload [16]. Among mental workload assessment methods, physiological metrics have gained popularity in the field of human factors research as a means of measuring mental workload levels, along with the importance and widespread use of subjective measures, job and task analysis, and performance assessments [17]. Physiological metrics offer valuable insights into how the human body responds to stimuli like work tasks. One benefit of psychological metrics is that they continuously and unbrokenly record data in real time [18]. However, interpreting these data is often challenging due to individual uniqueness. To address this, combining data with context or utilizing artificial intelligence analysis becomes necessary. Cardiac activity was handled by [19,20] to study the dynamics of mental workload. Exploiting brain activity, ref. [21] studied the subtleties of mental workload. In a different direction, ref. [22] focused on eye-tracking technology and investigated its uses to comprehend mental workload during HRC. Additionally, ref. [23] investigated the cognitive components of HRC using pupillometry. Significantly, refs. [4,24] adopted an all-encompassing strategy in their investigations, incorporating both cardiac and electrodermal activity-based sensors. Parallel to this, ref. [25] broadened the scope of their research by including sensors for cardiac, respiratory, and brain activity. The result was the creation of an adaptive human sensor framework specifically designed to improve human–robot collaboration. In an analogous setting, ref. [26] concentrated their research on body temperature, electrodermal activity, and cardiac activity, providing more insights into the dynamic, complex nature of physiological reactions in many contexts of HRC.

However, a consequence of the use of different physiological metrics to assess workload—both in isolation and in combination—in HRC is that it can be easy to lose track of which methodologies have been employed to measure different parameters. Thus, in this systematic review, we analyzed the different approaches used in the assessment of mental workload in HRC settings. We aim to provide a detailed review of methods and methodologies used to operationalize mental workload and potential applications for researchers and professionals. To achieve this, this review explores four key research questions:

- RQ1: What industrial tasks are employed to assess mental workload in HRC scenarios?
- RQ2: How are subjective and performance-based measures integrated with physiological data to assess mental workload in HRC?
- RQ3: What are the main physiological measures used to operationalize mental workload in HRC scenarios?
- RQ4: What are the challenges and limitations associated with using physiological measures to assess mental workload in HRC?

Several systematic reviews have explored the application of physiological sensors across various domains. Understanding and choosing suitable physiological metrics for mental workload (MWL) assessment in a variety of human–machine systems was presented by [27]. The research on in-vehicle psychophysiological measures to promote the creation of efficient driver support systems and human–machine driving interfaces was summarized by [28]. A wide variety of applied and experimental investigations from different domains were covered by [29], with safety-critical applications making up a large portion of the sample of applied literature evaluated. To the best of our knowledge, this is the first review that concentrates on the use of physiological sensors in the field of collaborative robotics in industry scenarios, particularly in tasks that require close cooperation between humans and robots. By addressing these research questions, this review aims to bridge the gap between mental workload assessment and HRC optimization, fostering the development of more intuitive, adaptive, and human-centered robotic systems.

## 2. Materials and Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist guidelines were applied to perform this systematic review. The PRISMA 2020 update provides new recommendations and explanations on existing ones to guarantee high-quality reporting [30]. The protocol was registered at PROSPERO International Prospective Register of Systematic Reviews of the University of York, with the registered number CRD42024510060.

### 2.1. Literature Search Strategy and Article Selection

An initial database search was conducted on Scopus without any time restrictions. The keywords used in this analysis were initially chosen through a preliminary review of several articles on the subject matter concerning the role of cognitive ergonomics in industrial HRC. Searches were conducted by 13 February 2025.

Considering three different topics (Human–Robot Collaboration, Cognitive ergonomics, and Psychophysiology), the search strings for each topic were established:

- Subject 1—HRC: Keywords: "collaborative robot*" OR cobot* OR "human robot collaboration" OR "human robot interaction";
- Subject 2—Cognitive ergonomics: Keywords: "mental workload" OR cognitive OR "mental fatigue" OR "mental effort";
- Subject 3—signal* OR biosignal* OR physiol* OR psychophysiol* OR "brain activity" OR eda OR "electrodermal activity" OR gsr OR "galvanic skin response" OR ecg OR electrocardiogra* OR ppg OR photoplethysmogra* OR fnirs OR "functional near-infrared spectroscopy" OR eeg OR electroencephalogra* OR respiratory OR "body temperature".

To conduct the search procedure in the scientific database, all conceivable combinations of these keywords—covering the titles, keywords, and abstracts of papers—were combined (using the Boolean 'AND' operator between keywords of each group and the 'OR' operator within each group).

### 2.2. Eligibility Criteria

The selection process involved several stages. First, duplicated papers were excluded. Second, the following eligibility criteria were considered: (a) full-length articles reporting empirical data and published in scientific journals, conference papers, and book chapters, thus excluding other works such as reviews, conference reviews, books, editorials, and surveys, and (b) papers not published in the English language were excluded. Searches were carried out without any timeframe constraints. In the following stage, papers with no full text available through the b-on consortium were excluded. Later, titles and abstracts were initially screened, and articles were excluded in three consecutive phases. First, exclusion was based on three specific criteria: (a) lack of relevance to the topic of mental workload, (b) focus on healthcare robots or health conditions, and (c) papers referred to social or cognitive robots. Next, a second screening was conducted to define the application scenarios, leading to the exclusion of studies related to virtual reality, teleoperations, search and rescue missions, unmanned aerial vehicles, extreme situations, and other scenarios not related to HRC in industry. After the authors completed their ratings, they gathered to finalize the list of which records to ultimately exclude and why. There was almost perfect agreement between the three raters in the assessment of the studies for exclusion based on their titles and abstracts. The inter-rater reliability, measured using the Kappa coefficient of Cohen, was 0.82 [31].

The full content of all remaining papers was read after the initial screening by a single author (E.P.) to determine which ones, in light of the previously stated eligibility criteria,

were the most pertinent. In this phase, papers lacking relevance to the application of physiological signals used to assess mental workload or containing marginal information were excluded from consideration. Finally, an analysis of the screened articles was conducted to identify different workload measurement types—namely performance, subjective, and objective measures—and their corresponding key indicators and specific features analyzed from signals to assess mental workload in HRC. Furthermore, data were also collected regarding the sample size employed in their studies, the type of tasks that were conducted, and the industry scenarios in which they can be categorized, as per the perspective of the authors of this review. At the end of this selection process, 25 papers were retained and deeply analyzed.

Figure 1 presents the PRISMA flow diagram representing a schematic illustration of the strategy adopted, displaying the number of studies included at each phase of the selection process and the reasoning for the exclusion.

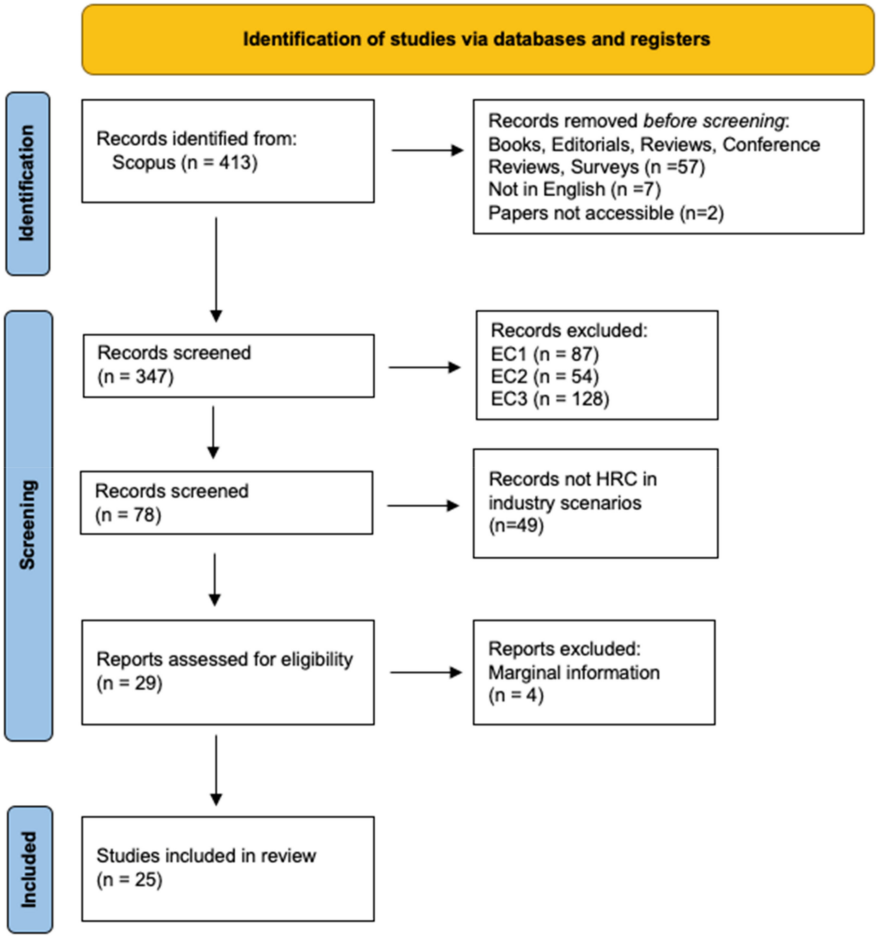


**Figure 1.** PRISMA 2020 flow diagram for the systematic literature review (EC: exclusion criteria). Adapted from [30].

## 3. Results

### 3.1. Main Findings

The systematic literature review identified 25 studies investigating mental workload in HRC across various industrial applications. These studies range from early implementations of machine learning for adaptive collaboration to recent advancements in neuro-ergonomic assessments, multimodal workload measurement, and robot adaptation strategies. Table 1 presents a chronological overview of the reviewed studies, a summary of their objectives from the perspective of the authors of this review, and information regard-

ing the publication type of each entry. When the publication year is considered, 6 papers were published between 2020 and 2022, 7 papers were published in 2023, 10 papers were published in 2024, and 2 papers were published in 2025. As for their publication types, 18 were published in scientific journals, while 7 were published in conference proceedings.

**Table 1.** Summary of studies on mental workload in HRC.

| Reference | Publication | Sample Size | Task Category | Workload Measure | Collected Data |
|---|---|---|---|---|---|
| [26] | CP | 4 | Assembly | Performance | Quality checks, quality issues, productivity, operating costs, variability of job, risk of accidents |
| | | | | Subjective | NASA-TLX; Borg CR-10; Job engagement |
| | | | | Physiological | HR; HRV; EDA |
| [32] | JP | 14 | Construction | Subjective | NASA-TLX; RS9 |
| | | | | Physiological | EEG |
| [33] | CP | 6 | Construction | Physiological | EEG |
| [34] | JP | 13 | Assembly | Subjective | NASA-TLX |
| | | | | Physiological | Cardiac Measures; EDA; Hand temperature |
| [35] | JP | 13 | Construction | Physiological | EEG |
| [36] | JP | 2 | Assembly | Physiological | EEG; EMG |
| [37] | JP | 48 | SIMKAP | Physiological | EEG |
| [38] | JP | 9 | Assembly | Performance | Number of errors/mistakes |
| | | | | Subjective | NASA-TLX |
| | | | | Physiological | EEG |
| [4] | JP | 36 | Assembly | Performance | Number of errors/mistakes |
| | | | | Subjective | NASA-TLX; SAM; BDM |
| | | | | Physiological | HRV (RMSSD); EDA (SCR); |
| [39] | JP | 14 | Construction | Subjective | NASA-TLX |
| | | | | Physiological | EEG |
| [40] | JP | 15 | Assembly | Performance | Number of errors/mistakes; time on task |
| | | | | Subjective | NASA-TLX; TAM3; Ad hoc acceptance, well-being and working experience; social impact |
| | | | | Physiological | OM; HR; video record |
| [41] | CP | 10 | Assembly | Physiological | EEG; EMG |
| [42] | JP | 13 | Pick-and-place | Performance | Number of errors/mistakes; reaction time |
| | | | | Subjective | NASA-TLX |
| | | | | Physiological | EEG fNIRS (HbO/HbR) |
| [43] | JP | 18 | Assembly | Subjective | Unstructured feedback |
| | | | | Physiological | EDA (SCL; SCR); HRV (RMSSD and SDNN); OM |
| [44] | CP | 15 | Material handling | Subjective | NASA-TLX |
| | | | | Physiological | fNIRS; EMG |
| [12] | CP | 12 | Assembly | Performance | Number of errors/mistakes |
| | | | | Subjective | Unstructured feedback |
| | | | | Physiological | EDA (SCR; SCL); HR |
| [8] | JP | 12 | Assembly | Performance | Number of errors/mistakes |
| | | | | Physiological | EDA (SCR; SCL); HRV (RMSSD) |
| [45] | JP | 6 | Assembly | Subjective | NASA-TLX; |
| | | | | Physiological | OM |

**Table 1.** *Cont.*

| Reference | Publication | Sample Size | Task Category | Workload Measure | Collected Data |
|---|---|---|---|---|---|
| [46] | JP | 24 | Blasting | Performance | Time-based measures |
| | | | | Subjective | NASA-TLX |
| | | | | Physiological | EEG (PEN, Pe) |
| [47] | JP | 4 | Assembly | Performance | Number of errors/mistakes; Time-based measures |
| | | | | Subjective | NASA-TLX |
| | | | | Physiological | EEG |
| [48] | JP | 22 | Inspection | Performance | Number of errors/mistakes |
| | | | | Subjective | NASA-TLX |
| | | | | Physiological | Cardiac activity; EDA; Temperature; ACC |
| [49] | CP | 11 | Problem-solving | Performance | Number of errors/mistakes |
| | | | | Subjective | NASA-TLX; TRUST |
| | | | | Physiological | EEG |
| [50] | CP | 43 | Assembly | Performance | Number of errors/mistakes |
| | | | | Subjective | NASA-TLX; TAM; Perceived control |
| | | | | Physiological | HR |
| [51] | JP | 17 | Inspection | Subjective | NASA-TLX; BWS |
| | | | | Physiological | HR |
| [52] | JP | 16 | Agricultural tasks | Subjective | NASA-TLX; NARS; RAS |
| | | | | Physiological | OM; Spine kinematics |

Moreover, it is observed that, regarding trends over the years, the data do not indicate consolidated patterns, showing variations in the combination of different physiological measures, whether applied in isolation or in conjunction with subjective and performance metrics, without a consistent adoption of approaches that reflect the researchers' perspectives.

However, the widespread adoption of subjective measures stands out, as they were used in 19 out of the 25 analyzed studies, with the integration of physiological and subjective measures establishing itself as the dominant and transversal approach in this review. Another relevant aspect is that performance measures, when applied, are always associated with subjective measures. In 11 studies, the joint application of all three cognitive workload assessment approaches was observed.

A notable aspect is the almost exclusive use of central nervous system-related metrics between 2020 and 2023, reflecting researchers' strong interest in analyzing brain activity and its behavior concerning workload in HRC settings. Additionally, central nervous measures were employed in 13 out of the 25 studies, followed by cardiac measurements and EDA, used in 10 and 7 studies, respectively. Ocular measures, mentioned in only four studies, were only observed starting in 2023. Finally, among the physiological metrics analyzed, temperature was employed by only two authors.

### 3.2. Characteristics of the Selected Studies

Using VOSviewer, authors' keyword co-occurrence network maps were generated to visually explore key research areas [53]. To ensure consistency in keyword representation, variations in formatting—such as acronyms, capitalization, and spacing—were standardized. For instance, occurrences of human–robot collaboration and EEG in different formats were uniformly converted to human–robot collaboration and electroencephalogram, respectively.

It is important to note that four papers did not provide author-defined keywords, and as a result, they were not included in this visualization. The network visualization of this bibliometric map, derived from the available bibliographic data, is presented in Figure 2.

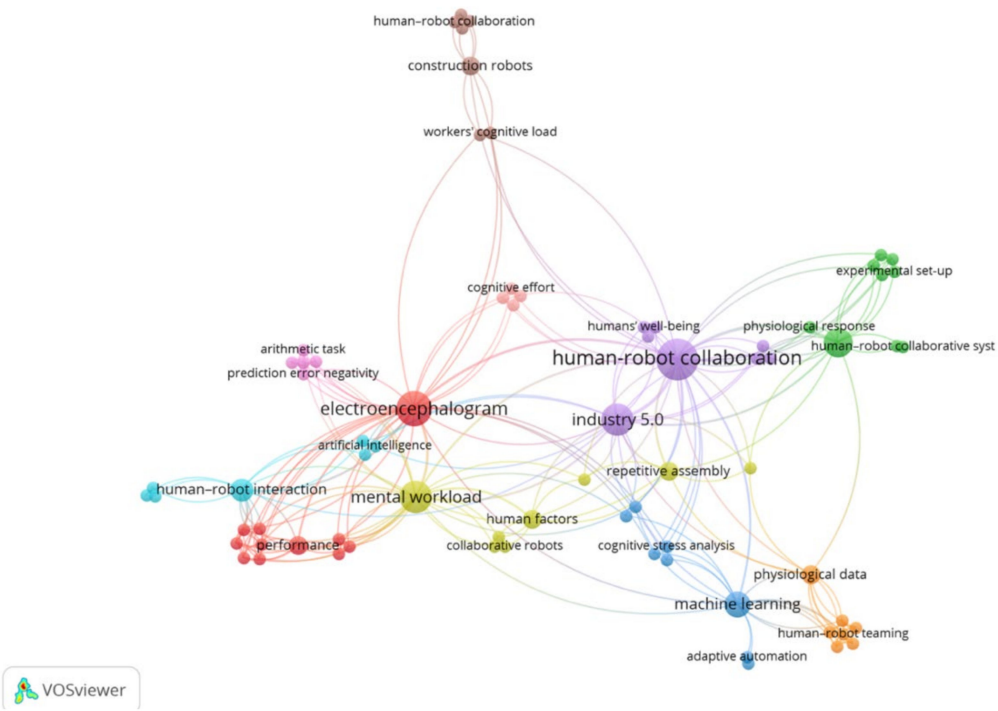


**Figure 2.** Illustration of most frequently author's keywords using VOSviewer software (v1.6.19).

In Figure 2, keywords are identified by a label inside a circle, with two items being connected by lines between their respective circles. The weight of the keyword determines the size of its respective label and circle, with larger labels and circles corresponding to heavier keywords, e.g., keywords with higher reoccurrence amongst papers. Furthermore, in terms of co-occurrence linkages, the proximity between two items indicates their correlation with each other: the stronger the relation, the closer two keywords are to one another. Analyzing the total incidence number at once reveals the correlation between all the keywords found in the papers.

The keyword co-occurrence network provides an insightful visualization of the key themes in human–robot collaboration (HRC) research, with a strong emphasis on HRC and electroencephalogram (EEG). At the center of the network, HRC emerges as the dominant theme, illustrating the growing interest in understanding human–robot interaction in industrial environments. The term Industry 5.0 is closely linked to HRC, reflecting the broader industrial context in which these studies are conducted. This is particularly relevant as Industry 5.0 emphasizes human-centric automation, fostering collaboration between robots and humans in ways that prioritize adaptability, well-being, and productivity.

Additionally, EEG is highly correlated with mental workload, indicating that cognitive workload assessment through physiological measures plays a fundamental role in recent studies. EEG provides real-time insights into cognitive states, allowing for adaptive systems that adjust robotic behavior based on human workload levels.

Clusters of interconnected keywords reveal distinct research directions. For instance, the cluster around electroencephalogram and mental workload connects strongly with terms such as prediction error and cognitive effort, indicating a focus on neurophysiological assessments of cognitive load during human–robot interactions. Another cluster centers around machine learning, physiological data, and adaptation, suggesting an interest in leveraging artificial intelligence techniques to enhance robot adaptability in collaborative environments.

Research on ergonomics and worker well-being appears in proximity to HRC, highlighting a growing concern with the physical and cognitive impact of robotic systems on

human operators. The presence of construction robots and assembly within the network suggests that studies are being applied to diverse industrial domains, ranging from smart factories to construction sites.

Overall, this keyword network highlights the interdisciplinary nature of modern HRC research, bridging cognitive science, physiological monitoring, and intelligent automation to create safer, more efficient, and human-centered collaborative robotic systems.

### 3.3. Operationalizing Mental Workload

A comprehensive overview of collaborative scenarios developed to analyze task distribution, movement patterns, and mental workload in HRC is available as Supplementary Materials (Table S1).

The reviewed studies span a wide range of industry scenarios, including manufacturing (10 papers), construction (4 papers), and, generically, the possible application in Industry 5.0 (11 papers).

Regarding collaborative tasks, the studies covered diverse activities, namely assembly (13 papers), blasting (1 paper), material handling (1 paper), quality inspection (2 papers), problem-solving (1 paper), construction-related tasks (4 papers), agricultural tasks (1 paper), pick-and-place (1 paper), and simultaneous capacity multitasking (SIMKAP, 1 paper). Common specific tasks involve assembling electronic devices, mechanical systems, and industrial prototypes, as well as transporting objects, bricklaying, and quality control inspections. These tasks were often performed under different experimental conditions to evaluate human cognitive load, efficiency, and task adaptability when working with collaborative robots.

Frequent movement patterns observed in HRC scenarios include object manipulation (e.g., assembly, pick-and-place), coordination with robot movements (e.g., synchronized handling and delivery of components), and real-time adjustments based on mental workload feedback. Key aspects of these patterns involve precise hand–eye coordination, timing, and task sequencing to maintain efficiency and performance. In some scenarios, robots dynamically adjusted their behavior by modifying speed, providing cognitive support, or optimizing task allocation.

One key difference across studies is the sample size, which varied significantly. Ten to twenty participants were included in most of the research sample sizes. Some studies involved a small number of participants, such as 2 participants [36] or 4 participants [26,47], while others included larger samples, such as 34 [50], 36 [4], and 48 participants [37].

### 3.4. Performance Assessment of Mental Workload

The two most common types of performance measures used across the reviewed studies were the number of errors/mistakes during tasks [40,48–50] and time-based performance measures, such as time to complete tasks [40,47] and reaction time [42].

### 3.5. Subjective Assessment of Mental Workload

Considering Figure 3, the NASA-TLX questionnaire stands out as the most widely used subjective workload measure, appearing in 61% of the studies. In addition to NASA-TLX, the Bedford Workload Scale—another NASA-developed scale—was also utilized [51]. Unlike multi-dimensional rating scales such as NASA-TLX, the Bedford scale is unidimensional and specifically designed to assess an operator's spare mental capacity.

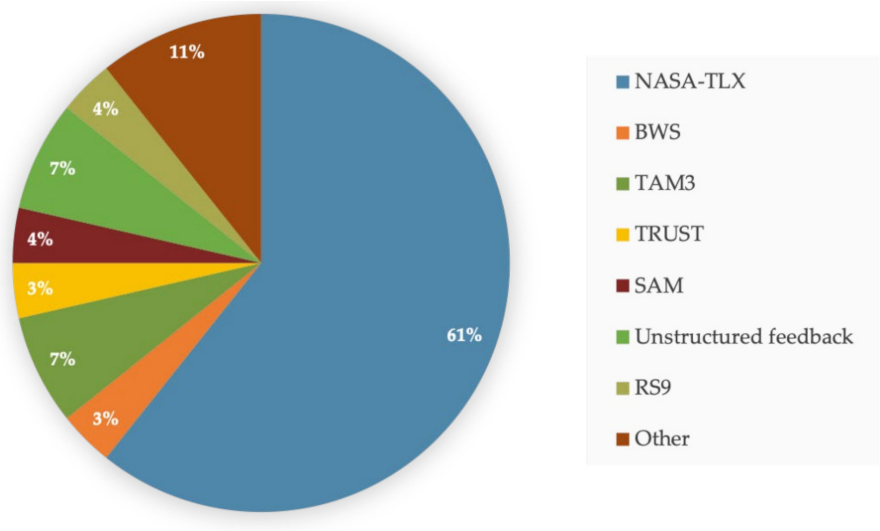


**Figure 3.** Workload assessment through subjective measures.

Beyond those measures, several researchers have employed other subjective metrics to evaluate factors associated with mental workload. For instance, refs. [4,43] gathered qualitative insights through unstructured feedback. Emotional responses were analyzed by [4] using the Self-Assessment Manikin (SAM). Meanwhile, ref. [40] combined NASA-TLX with task-specific scales, such as the Technology Acceptance Model (TAM), along with customized well-being and work experience surveys.

### *3.6. Physiological Assessment of Mental Workload*

Among the most common physiological measures are central nervous measures and peripherical parameters, such as cardiac measures, electrodermal activity, ocular measures, and hand temperature.

#### 3.6.1. Central Nervous Measures

Central nervous measures have been predominantly used, observed in 13 of the reviewed studies, underscoring the centrality of neurophysiological metrics to capture real-time brain activity and assess mental workload levels during collaborative interactions.

Depending on their intended use, neuroimaging methods can be divided into two primary categories: those that aim to measure neural activity indirectly by using metabolic indicators, like fNIRS, to provide a comprehensive picture of cognitive states and those that measure neural activity directly in response to stimuli using an EEG [54].

##### Electroencephalogram

The historical trajectory of understanding the brain's electrical activity, initiated by Richard Caton in 1875, has evolved with contemporary technologies like the electroencephalogram. This non-invasive method captures rapid changes in the brain's electrical signals resulting from ion movements within neurons [37,55]. Employing multiple electrodes on specific scalp regions, often following the 10–20 system [56], EEG unveils standardized rhythmic patterns in alpha (7–12 Hz), theta (4–7 Hz), beta (13–30 Hz), delta (0.5–4 Hz), and gamma (>30 Hz) frequency bands. These frequency bands correspond to distinct mental processes, such as relaxation, focus, executive control, and cognitive stress. Delta frequency is associated with deep sleep [37,38].

EEG spectral power analysis has shown consistent patterns of activation and inhibition during HRC tasks, particularly in response to cognitive workload fluctuations. The results indicate a decrease in parietal alpha power as cognitive load increases [36,46], supporting

the notion that alpha activity is linked to relaxation and reduced engagement [42]. Additionally, high-alpha power was more prominent in the maladaptive condition, possibly reflecting an increased effort to inhibit irrelevant information due to robotic errors [49]. Beta power increased in response to heightened cognitive load and task complexity, indicating greater mental effort and alertness during demanding tasks [42]. Theta power increased significantly in the frontal region following unexpected robot stops, particularly under high cognitive load conditions, suggesting heightened working memory demands and executive control during error detection [46]. Additionally, higher theta power was observed in the central and temporal regions when participants performed tasks alone, indicating a greater focus on internal task representation, while in the maladaptive condition (inexperienced robot), theta power increased in the occipital region, reflecting greater visual processing due to robot errors [49]. Furthermore, an increase in frontal theta activity was linked to mental effort, reinforcing its role in working memory activation and executive control [36].

To assess cognitive load, researchers employ EEG-derived ratios and indices that provide insights into attentional focus, mental workload, and cognitive engagement during task execution (Table 2).

**Table 2.** Power ratio interpretation (adapted from [41] and expanded by the authors).

| Spectral Power Band | Interpretation | Employed in |
|---|---|---|
| Alpha | Linked to relaxation and idle mental states. Alpha power decreases as cognitive demand increases. | [32,36,42,46,49] |
| Theta | Related to cognitive control, working memory, and sustained attention. Frontal theta power increases with higher cognitive demand. | [36,37,46,49] |
| Betha | Associated with active cognitive processing, stress, and alertness. Beta power increases in response to heightened cognitive load and task complexity. | [32,38] |
| Gamma | Linked to complex cognitive functions, memory processing, and high attentional states. Gamma oscillations correlate with high mental effort and concentration. | [32,37,41] |
| Alpha/Beta | Higher Alpha/Beta ratio = relaxed mind<br>Lower Alpha/Beta ratio = alert state of mind | [37,41] |
| Beta/Alpha | Higher Beta/Alpha ratio = alert state of mind<br>Lower Beta/Alpha ratio = relaxed mind | [38] |
| Alpha/theta | Higher Alpha/Theta = focused and alert state of mind<br>Lower Alpha/Theta = more relaxed and meditative states | [37,41] |
| Theta/Alpha | Higher Theta/Alpha = more relaxed and meditative states<br>Lower Theta/Alpha = focused and alert state of mind | [36,47] |
| Theta-Beta ratio | Associated with working memory and attentional control. | [47] |
| Gamma/Theta | Research has shown that the gamma/theta ratio is higher during states of focused attention, such as when performing a visual or auditory task.<br>The Gamma/Theta ratio has also been linked to memory processing, with higher ratios observed during successful encoding and retrieval of memories. | [37,41] |
| Beta/Alpha + Theta | Reflects mental effort, vigilance, attention, alertness, and task engagement. | [47] |

The Alpha/Theta ratio ($\alpha/\theta$) is commonly used to measure attentional focus, with higher values indicating increased cognitive engagement. In experimental conditions, an increase in this ratio suggested enhanced focus and attention during task execution [37]. The Alpha/Beta ratio ($\alpha/\beta$), associated with relaxation versus cognitive effort, showed an increase post-task, reflecting reduced stress and a transition to a more relaxed state [37,41].

Conversely, the Beta/Alpha ratio ($\beta/\alpha$), which is indicative of mental workload, demonstrated distinct patterns. In standard conditions (without robotic assistance), it remained stable or increased slightly throughout the task, whereas in a collaborative robot-assisted scenario, a significant reduction was observed, particularly in later task

phases, suggesting a decrease in cognitive load over time [38]. The Gamma/Theta ratio ($\gamma/\theta$), associated with working memory and concentration, exhibited peaks during the most cognitively demanding task phases, particularly during the initial assembly and repositioning of the plate. This pattern indicates heightened cognitive demands, increased attentional focus, and active memory processing [37,41].

The Theta/Alpha ratio ($\theta/\alpha$), an indicator of cognitive effort, exhibited an increase as workload levels intensified [47]. In collaborative settings, the presence of a robot significantly reduced $\theta/\alpha$, suggesting a lower mental workload. This reduction was further supported by its strong correlation with mental demand scores in the NASA-TLX questionnaire [47]. Additionally, the Mental Workload Index proposed by [36], which is calculated as the ratio of frontal theta power (Fz) to parietal alpha power (Pz), peaked during the most cognitively demanding phases of the assembly task. Similarly, ref. [41] observed that the Alpha/Theta ratio increased during assembly, indicating a heightened state of focus and attention.

The Theta/Beta ratio ($\theta/\beta$), associated with working memory and attentional control, showed that an increase in this index may signal reduced attentional control in high-demand tasks [47].

Lastly, besides extracting information from frequency domain bands, ref. [32] also extracted information from the time domain, namely mean, peak location, and kurtosis. Ref. [37] analyzed the variations in amplitude and fluctuations over time under different workload conditions. Ref. [46], in turn, used error-related potential (ERP) to detect errors. ERP consists of a time-locked negative peak, or error-related negativity, around 200 ms after the error occurrence. It was mostly followed by a positive peak around 300 ms after the error onset, and this error positivity (Pe) is related to the user's error awareness. In their study, prediction error negativity (PEN) was measured as the maximum error negativity amplitude in the 100–350 ms window. In settings with active movement of participants, the prediction error positivity (Pe) was measured as the maximum error positivity in the 350–450 ms window. PEN was observed during the execution of a motor task prior to any external error indication. Pe was used to study error awareness, and the study concluded that increased workload depreciates the human awareness of the human operator. The study used error-related potentials (ERP), specifically prediction error negativity (PEN) and error positivity (Pe), to evaluate error awareness during physical human–robot collaboration. PEN, a negative peak around 200 ms after an error occurrence, and Pe, a positive peak around 300 ms, were analyzed under different levels of mental workload. The results showed that the amplitude of PEN and Pe decreased with increasing mental workload, indicating reduced error awareness.

Regarding EEG limitations, EEG devices are considered highly sensitive instruments that frequently capture substantial levels of noise, which can be categorized into intrinsic (originating from physiological processes such as facial movements, body motion, and ocular activity) and extrinsic (stemming from environmental or physical sources, such as the existence of background noise or electrical currents) [32,37]. This noise significantly compromises the quality of EEG recordings. Moreover, EEG signals are non-stationary, which can affect classifier accuracy over time, making it challenging to generalize cognitive workload prediction models [33]. Another critical limitation is the variability in neurocognitive responses among participants, which may influence data interpretation and the reliability of inferred conclusions [36,49]. Additionally, participants' physical characteristics, such as thick or curly hair, can hinder proper electrode and optode contact, compromising data quality [42]. Lastly, inconsistencies in electrode placement during data collection can introduce signal variations, reducing the robustness of EEG-based models [37].

Functional Near-Infrared Spectroscopy

Functional near-infrared spectroscopy (fNIRS) is a non-invasive neuroimaging technique that measures the hemodynamic response of the brain. This technique measures changes in oxy- and deoxy-hemoglobin concentration in specific brain regions in contrast to a baseline period. fNIRS detects the hemodynamic changes in response to a cognitive stimulus, allowing the localization of brain responses to specific cortical regions [57]. To assess mental workload, this technique may allow researchers to verify if there is an increase in neural efficiency and functional connectivity in frontoparietal networks, which are involved in executive function and working memory workload [58]. Increased levels of oxy-hemoglobin and decreased deoxy-hemoglobin concentrations, in contrast to baseline measurements, are indicative of significant brain activations. Reduced brain oxygenation is indicative of fatigue, whereas elevated cerebral oxygenation is linked to mental stress and workload [28]. According to [44], HRC resulted in a decrease in oxy-hemoglobin (i.e., reduced activation) in the prefrontal cortex when compared to human–human collaboration and solo work. This led the authors to suggest that HRC may have reduced the participant's mental workload.

However, one significant drawback of fNIRS is that its temporal resolution is constrained by the time course of hemodynamic activity (on the order of seconds), as it depends on measuring the absorption characteristics of light as a function of vascular alterations in the brain. To facilitate real-world monitoring in ecologically sound environments, recent work has proposed the simultaneous collection of fNIRS and EEG signals [57,59,60]. Ref. [42] combined EEG and fNIRS data to analyze how task complexity, robot speed, and robot payload capacity affect the perceptual state of factory workers under cognitive load conditions by monitoring this impact in subjective, behavioral, and physiological measures. The left prefrontal cortex exhibits a comparatively larger concentration of oxy-hemoglobin, particularly during high-complexity episodes. This multimodal approach improved prediction accuracy by capturing a broader spectrum of physiological responses [61].

### 3.6.2. Cardiac Measures

Cardiac measures, featured in 10 of the reviewed studies, also stand as a significant contributor, affirming the physiological responsiveness of the cardiovascular system as a valuable measure for workload assessment.

Heart rate (HR), defined as the number of heartbeats per minute (bpm), has traditionally been the most used signal feature. However, findings on its sensitivity to workload variations in HRC remain mixed. Several studies reported no significant differences in HR across experimental conditions.

Ref. [12] found that the average heart rate did not significantly differ between HRC and manual task scenarios. Similarly, ref. [40] observed no significant variations in HR, reinforcing the idea that HR alone may not always be a strong discriminator of cognitive load in collaborative environments. Ref. [50] further confirmed this, reporting no significant difference in subjective mental workload or HR between adaptive and standard speed conditions in an industrial HRC set-up. On the other hand, some studies have indicated a moderate correlation between HR and cognitive load. Ref. [34], in their study on classifying cognitive load in HRC using physiological feedback, identified HR as sensitive to changes in mental effort. Supporting this, ref. [51] found a moderate positive correlation between cognitive load and HR, suggesting a slight increase in physiological activation with higher mental workload in collaborative industrial systems.

In contrast, heart rate variability (HRV), the fluctuation in the time intervals between contiguous heartbeats, has, in turn, been touted as the most promising indicator for describing workload [18]. Time- and frequency-domain and nonlinear analysis are thought of

as the main divisions of cardiac measures analysis, reflecting distinct aspects of autonomic nervous system modulation.

Within the time-domain analysis (Table 3), the most relevant HRV features include SDNN (standard deviation of all successive normal R-R intervals), RMSSD (root mean square of the successive R-R interval differences), PNN50 (proportion of beats with a successive R-R interval difference exceeding 50 ms), and HRVTi (heart rate variability time index). While HRVTi and SDNN are associated with general autonomic influence, PNN50 and RMSSD are more directly linked to parasympathetic (vagal) activity [62].

**Table 3.** Common HRV features in the time domain.

| Features | Description | Employed in |
|---|---|---|
| NN | Normal-to-normal interval. Also called the R-R interval or the interbeat interval (IBI). Measures the time between QRS peaks. | [8] |
| SDNN | The standard deviation of all NN intervals. | [43] |
| SDANN | The standard deviation of the averages of NN intervals in all 5 min segments of the entire recording. | -- |
| RMSSD | The square root of the mean of the sum of the squares of the differences between adjacent NN intervals. | [4,8,43] |
| pNN50 | Proportion of differences in consecutive NN intervals that are longer than 50 ms. | -- |
| HRVTi | The sum of all R-R intervals divided by the maximum density distribution. | -- |

Note: -- not employed in any of the reviewed studies.

Despite HRV being widely used, its relationship with task complexity and robotic assistance remains unclear. Ref. [43] found that HRV, measured through RMSSD and SDNN, did not show significant variations between manual and collaborative conditions, suggesting that task modality alone may not be a determining factor for physiological workload variations. However, other studies have reported more specific trends. Ref. [4] found that RMSSD was significantly higher in the manual assembly condition at both mid-session and the end of the task, indicating lower physiological stress. Furthermore, RMSSD showed a general increase throughout the session, suggesting progressive adaptation to the task. Additional insights from ref. [8] suggest that RMSSD was consistently lower in the HRC setting, implying that the operator maintained a sustained level of attention, possibly due to the structured nature of robotic assistance. In contrast, higher RMSSD values in manual conditions may indicate moments of relaxation or cognitive disengagement, potentially reflecting a greater likelihood of distractions in non-automated environments.

The distribution of absolute or relative spectral power into four frequency bands is estimated by frequency-domain features (Table 4) named ultra-low-frequency, very-low-frequency, low-frequency, and high-frequency bands [63]. The power spectral of each frequency component, as well as the total power spectral of all frequency bands, are thought to be useful features for measuring mental workload. Sympathetic activity is assumed to be associated with LF-relative features. It is believed that there is a correlation between parasympathetic activity and HF-related features. With long-period rhythms, the physiological significance of the VLF-related features has been determined [62]. It has been demonstrated that the LF/HF ratio reflects the balance between the parasympathetic nervous system and the sympathetic nervous system, also known as the "fight or flight" system. Because of this, when determining how sympathetic the response is in relation to parasympathetic, the LF/HF ratio is frequently used. Moreover, frequency-domain HRV data have been employed owing to their documented association with mental workload,

characterized notably by a decrease in parasympathetic activity or an increase in LF/HF ratio [34].

**Table 4.** Common HRV features in the frequency domain.

| Features | Description | Employed in |
|---|---|---|
| Ultra-Low-Frequency (ULF) | Power spectrum $\leq$ 0.003 Hz | -- |
| Very-Low-Frequency (VLF) | Power spectrum from 0.003–0.04 Hz | -- |
| Low-Frequency (LF) | Power spectrum from 0.04 to 0.15 Hz | [34] |
| High-Frequency (HF) | Power spectrum from 0.15 to 0.4 Hz | [34] |
| Sympathetic Modulation Index (SMI) | SMI = LF/(LF + HF) | -- |
| Vagal Modulation Index (VMI) | VMI = HF/(LF + HF) | -- |
| Symphatovagal Balance Index (SVI) | SVI = LF/HF | [34] |

Note: -- not employed in any of the reviewed studies.

Lastly, the unpredictability of a time series can be measured using nonlinear features, as the time-domain and frequency-domain characteristics of HRV signals fall short of fully capturing the nonlinear aspects of HRV signals. Some authors are currently using nonlinear analysis methods to analyze HRV signals due to mental workload, including sample entropy (SaEn) and detrended fluctuation analysis (DFA) [62].

Blood pressure measurements, such as systolic, diastolic, and mean arterial pressure, were frequently employed to measure MWL in addition to ECG measurements [34,48]. Blood pressure (BP) is a measurement of the force that blood flowing through the body exerts on the blood vessel walls. The pressure that is applied as the heart muscle contracts (the systole) and relaxes (the diastole) is the standard way that blood pressure is expressed. Usually, an increase in systolic BP with workload and stress is observed.

Peripheral oxygen saturation (SPO2) is a measure that indicates the percentage of hemoglobin in the blood that is saturated with oxygen. Hemoglobin is a protein found in red blood cells that transports oxygen from the lungs to the body's tissues. $SPO_2$ is measured non-invasively, typically using a device called a pulse oximeter, which is placed on the finger, earlobe, or other parts of the body. SPO2 was used by [34] to discriminate different cognitive tasks.

Regarding its limitations, cardiac measurements rely on wearable devices, such as chest straps and smartwatches, for the collection of physiological data. While these devices are non-invasive, their accuracy is contingent upon the quality of the equipment and its application. Additionally, cardiac measures are highly susceptible to external factors, including workplace conditions such as noise and temperature, as well as the operator's emotional state, which may introduce inaccuracies—a challenge inherent to other physiological signals. Moreover, depending on the intervals used for data collection, the system may fail to capture rapid fluctuations in the operator's mental state, thereby limiting its ability to promptly detect and respond to sudden variations in workload levels.

#### 3.6.3. Ocular Measures

Recently, there has been a significant increase in the use of eye-tracking technology in the context of HCR [64], largely due to advancements that have made these tools more sophisticated and widely available. Eye tracking goes beyond simply noting where the eye is looking; it allows the detailed examination of where, how long, and in what sequence the eyes focus on different points in the visual field. It is a valuable tool used to analyze visual attention through fixations, saccades, and pupil dilation [40,45], offering insights

into how cognitive load affects visual processing and attention, with variations in these signals providing key indicators of mental workload fluctuations (Table 5).

**Table 5.** Common ocular measures and respective definitions (adapted from [45]).

| Features | Description | Employed in |
|---|---|---|
| Blink rate | Blink frequency per minute or second. Higher blink rates can be associated with higher mental demand or fatigue, while lower blink rates can be associated with higher visual demand or attention. | [40,52] |
| Blink duration | Closure time duration of a blink. Lower blink duration may be associated with higher visual demand, while higher blink duration can be provoked by tiredness or fatigue. | [40] |
| Pupil size | Diameter or area of the pupil. Pupil size in adults can range between 2 mm and 8 mm in diameter. Higher pupil size can be associated with higher mental demand. | [43,45,52] |
| Fixation rate | The number of fixations, usually in a certain area of interest (AOI). The number of fixations approximates visual attention allocation. More fixations can equate to less efficient search or increased visual effort, thus, a higher mental workload. | [40,43,45] |
| Fixation duration | The time spent gazing at a position. A longer fixation duration describes issues related to extracting information (i.e., more processing time), or it indicates that the target is more appealing. | [40,43,45] |
| Saccade rate | The number of saccades, usually in a certain AOI. A higher number of saccades can be associated with higher visual effort and, thus, higher mental workload. | [43,45] |
| Saccade duration | The length of time from the start to the end of a saccade event (i.e., shifting from one fixation to another). | [43] |
| Saccade amplitude | The measure of visual arc degrees of movement from one fixation to the next. Saccade amplitude usually drops as mental workload increases. | [43] |
| Saccade velocity | The speed of the saccade (degrees/time) is usually measured considering the peak velocity. | [43] |

Research conducted by [43,45] underscores the reliability of eye-tracking in long-duration sessions. Their findings indicate a positive relationship between pupil diameter (pupil size) and cognitive processes. Likewise, ref. [52] observed that pupil diameter was reduced in collaborative robot modes (both with and without cues), suggesting a lower mental workload.

In the same study, ref. [52] also noted that, in comparison to the manual mode, the number of blinks in the "unreliable condition" (participants were informed that the robot could make mistakes and they had to correct them) increased, which suggests an increase in the subject's cognitive load due to the need to supervise and correct the robot's mistakes. In a similar vein, ref. [40] also found that the number of blinks increased more when participants had to do a secondary task (do mathematical tasks aloud alongside the assembly task) in comparison to both the single task (assembly task) and the rest mode. Regarding blink duration, in turn, while the authors found no differences between the dual and single task conditions, in comparison with the rest mode condition, blink duration was longer on both tasks. The authors also found similar results regarding fixation duration as well.

As for saccade amplitude, this metric was only employed in [43]. In their study, the authors found that in a collaborative scenario, in comparison with the manual modality, saccade amplitude was higher, indicating a reduction in the subject's cognitive load. In that same study, the authors also analyzed the fixation rate and found that it increased

with more complex tasks. Similar results were found by [45], who reported that the rates of fixations and saccades increased as cognitive load increased, not only while participants learned how to perform an assembly task but also while they had to resume it as well. In contrast, both of these rates decreased throughout a shift as participants got used to the assembly task. However, a relation between increased fixation rate and increased mental workload was not reported by [40], who employed this metric as well.

Some contradictory results highlight how different tasks can have varying effects on physiological responses, an issue that should be considered in future investigations [40,64]. Moreover, limitations of ocular measures also include significant variability among individuals due to personal differences, such as ocular characteristics and levels of experience, which can impact the accuracy and consistency of the collected data [45].

#### 3.6.4. Electrodermal Activity

Electrodermal activity (EDA), also known as galvanic skin response, is a gold-standard bioindicator that is frequently employed to assess mental workload. The electrical potential created on the skin's surface due to sweat gland activity is measured by the EDA sensor [24]. The use of EDA to gather psychophysiological data is non-invasive, reliable, affordable, and simple [65,66]. Unlike other target organs of the human body, which are connected to the sympathetic and parasympathetic nervous systems, the skin and sweat glands are exclusively innervated by the sympathetic nervous system. This makes EDA an ideal indicator of sympathetic activation and, therefore, a good indicator of when a response is triggered due to mental workload [67].

Two components in the high-resolution EDA signal were found [24], i.e., the tonic (skin conductance level, SCL) and phasic response (skin conductance response, SCR), and these derived metrics were used to quantify cognitive states and stressful periods [68]. Skin conductance shows an overall rising tendency over time, especially in humid environments, referred to as the tonic skin response [69]. The tonic phase, which establishes the baseline skin conductance, evolves gradually and exhibits minor fluctuations in the range of 10 to 100 s. Depending on a subject's level of hydration, the condition of their skin, or their autonomic regulation, the signal's rise and fall fluctuates continuously within that subject. Additionally, among subjects, this response can differ greatly [70]. The palms and soles, which are areas where sweat glands are most active, are the optimal places to collect such signals [71], although it has been demonstrated that the shoulder is an excellent alternative spot to capture this signal [24].

Physiological indicators in EDA studies primarily centered around SCR, as it provides a reliable indicator of the body's answer to mental workload. Its capacity to record variations in sympathetic nervous system activity in response to workload fluctuations offers insightful data on the person's level of mental exertion while performing a task [72,73]. In the study conducted by [4], EDA was found to be sensitive to task complexity and robot support, with higher SCR and SCL values observed in more complex tasks and in manual mode. Moreover, according to ref. [43], EDA, measured through SCL and SCR, also increased with assembly complexity. In the collaborative setting, EDA values were generally lower, suggesting a reduction in mental workload when the robot was present.

Regarding the limitations of EDA, signal fluctuation varies continuously within the same individual due to factors such as hydration, dry skin, and autonomic regulation. Additionally, this response can differ significantly between individuals [70], as sweat gland activity is influenced by both internal and external factors. Internal factors include age, gender, and ethnicity, while external factors encompass temperature, humidity, time of day, season, and medication use [74].

#### 3.6.5. Temperature

Healthy body temperature is typically between 36.5° and 37.5° and depends on various factors, one of them being the location or body part where the temperature is being measured. Body temperature can also change with emotions, the state of consciousness, arousal, or, in other words, with the activity of the sympathetic nervous system, which can lead to vasoconstriction or vasodilatation, thus causing fluctuations in temperature.

The temperature that is being measured, depending on the method and location, is not necessarily the body temperature because it can be different across different skin areas and body cavities. Usually, rectal, buccal, or tympanic temperature measurements are considered the most representative of true body temperature. All these locations are intrusive and thus inadequate in industrial/laboratory experimental settings. For this reason, hand temperature has been used [34,48] and has shown a high correlation with mental workload [34].

## 4. Discussion

This systemic review focuses on the topic of how previous studies have assessed mental workload in HRC scenarios. To this end, we collected data from 25 papers of an initial pool of 413 papers. From these 25, information was extracted regarding which tasks they employed to induce mental workload in participants working with robots, as well as what physiological, subjective, and performance data were collected and reported in each paper. From the information collected from each of these papers, we also presented a brief summary regarding each of the metrics that were found to be employed throughout the 25 papers, including information regarding both their strengths and limitations.

The importance of mental workload in industries has recently gained recognition due to accumulated knowledge generated in recent years. The studies reviewed in this paper focused on tasks like assembling, blasting, material handling, quality inspection, problem-solving, construction-related tasks, agricultural tasks, pick-and-place, and simultaneous capacity multitasking. These tasks are inherently physically and mentally demanding and thus can benefit from precise assessments of mental workload to ensure a worker's optimal performance, efficiency, and, most importantly, safety. However, the complexity of the human factor, which encompasses multiple variables, as well as the inherent differences between different tasks and environments, among other factors, makes it difficult to develop a worldwide and standardized concept [10]. The diverse measures of mental workload and their interactions further complicate developing a unified idea and also make it challenging to compare among the different studies. The lack of a holistic concept makes it difficult to adopt tried-and-true tactics, which is an obstacle to the creation of management systems. Nonetheless, through the combined use of physiological, subjective, and performance-based measures, researchers are making strides in effectively operationalizing and managing mental workload.

Physiological sensors played a crucial role in this process, with brain measures being the most frequently used method (Table 1), as they allow for brain activity to be directly measured. However, despite promising results, defining the specific spectral frequency that best characterizes changes in mental workload, as well as the location of electrodes, remains divergent. Still, several studies have consistently reported increased theta and beta bands located in the frontal lobe (related to executive functions) during periods of elevated mental workload [46]. The alpha band, typically associated with relaxation, showed a significant decrease under increased cognitive load in several studies [36,42,46,49]. While EEG has been widely explored for workload assessment, its potential extends far beyond this application. Given its ability to capture real-time, objective data on cognitive and emotional states, researchers have increasingly investigated its use in other domains. For instance,

EEG is being integrated into user experience evaluation, providing valuable insights that complement traditional subjective measures. Recent studies have demonstrated that EEG can effectively reflect user satisfaction and emotional responses during human–robot interactions, offering a deeper understanding of both conscious and subconscious reactions that subjective metrics might fail to capture [75].

Currently, there is a growing trend toward the simultaneous collection of brain signals using EEG and fNIRS, as the two neuroimaging techniques complement each other with the temporal resolution of EEG and the spatial resolution of fNIRS [42]. However, even though optical imaging techniques have a higher spatial resolution than EEG, their ability to draw conclusions is constrained by the shallow penetration depth of NIR light, which is just a few centimeters below the scalp's surface. Consequently, there is limited ability to image the activity of deep cortical and subcortical sources (beyond the outer cortical mantle) [42,44].

Data from cardiac measures, in turn, are also commonly used to evaluate mental workload. HRV, for example, can be used as an indicator of workload [18], as changes to it can reflect responses from the autonomic nervous system [62,76]. For example, a decrease in HRV can indicate mental workload. Moreover, in the extraction of HRV features, the LF/HF ratio is frequently used to assess mental workload. As mental workload increases, parasympathetic activity decreases, and an increase in the LF/HF ratio is observed [76].

Lastly, feature selection in EDA studies is primarily centered around skin conductance response, as it provides a reliable indicator of the body's response to mental workload [4]. Its capacity to record variations in sympathetic nervous system activity in response to workload fluctuations offers insightful data on the person's level of mental workload while performing a task [72].

Although brain and cardiac measures are dominant, there is a clear trend toward the exploration of integrated variables to assess mental workload in HRC. The study conducted by [77] employed a deep neural network (DNN) model to classify mental workload levels based on EDA, EEG, and PPG signals, achieving a classification accuracy of 86% and highlighting the value of multimodal physiological signals for robust workload detection. The combination of ocular measures with heart rate in [43] to assess cognitive workload during assembly tasks further illustrated the utility of multimodal sensing in accurately assessing mental workload.

The increase in the use of more integrated methodologies points towards a more holistic approach to mental workload evaluation in recent years as researchers build upon prior studies. It is important to note that some researchers still advocate for the adoption of alternative or more tailored measures, suggesting that the current trend may not entirely reflect every researcher's perspective. For example, subjective measures have remained important throughout the research timeline, with tools like the NASA-TLX being consistently employed. While physiological measures capture real-time data on brain activity and performance metrics provide task-related outputs, subjective measures offer insights into how individuals perceive their workload. Furthermore, subjective metrics can also be applied to assessing a subject's user experience, affective states, and stress [78], which can also be useful metrics to evaluate other aspects of the human–robot relationship and collaboration. Nevertheless, they frequently overlook the nuances of unintentional emotional reactions or risk distortion brought on by subjective assessments. Thus, the use of facial expressions for detecting cognitive and emotional states has gained prominence in various research fields, such as psychology, marketing, medicine, and human–computer interaction [79]. In terms of performance metrics, key indicators such as task accuracy, reaction time, and error rates are consistently employed across studies to quantify mental workload. For instance, multitasking and safety-critical operations often utilize EEG and

HRV in tandem with performance measures to offer a more comprehensive assessment of mental workload [62,76]. The inclusion of these metrics in both controlled experiments and real-world settings underscores their reliability and relevance to managing workload in operational environments [80].

Despite the promising results of physiological metrics for the assessment of mental workload, several limitations need to be acknowledged. One primary limitation is the small sample size across most of the reviewed studies, which can significantly affect the generalizability of the findings. For instance, in [36], the authors used only two participants, limiting the robustness of the conclusions. Furthermore, controlled experimental settings may not fully capture the variability and unpredictability of real-world environments, particularly in high-stakes fields like construction and industrial settings [32,35]. Moreover, the instruments that allow this are not without their own limitations. For example, the efficiency of measuring brain data with tools such as the EEG is influenced by various factors, like intrinsic (e.g., facial movements and body motion) and extrinsic artifacts (e.g., background noise or electrical currents). Another critical limitation is the variability in physiological signal interpretation among different subjects. Variations in EEG signal patterns, HRV, and other physiological responses can lead to inconsistent workload assessments. This situation presents challenges to the development of universal methods or machine learning-based models, as noted in [42,46].

Technical challenges also present significant limitations. For example, the accuracy of EEG signal recognition is crucial for the system's effectiveness but can be hampered by noise and artifacts in the data [25,77]. Similarly, the integration of multimodal data (e.g., EEG, fNIRS, EDA) requires sophisticated processing techniques and reliable synchronization, which can be technically demanding and prone to errors [25,42]. Moreover, the reliance on specific machine learning models, while beneficial for analyzing physiological data, also has limitations. Models like SVM, KNN, and gradient boost (GB) require careful tuning and validation to avoid overfitting, particularly with small datasets. Therefore, further validation across different scenarios, including laboratory and real-world settings, is essential to guarantee the wider applicability of these models [37].

Future research should address these limitations by incorporating larger, more diverse sample sizes and validating findings in real-world environments, especially in scenarios that might not be well-addressed by current tools. There is also a need to refine machine learning algorithms to improve their generalizability across different tasks and settings. Exploring additional physiological indicators and integrating them with existing measures could provide a more comprehensive understanding of mental workload. Advancements in wearable technology and real-time data processing will further enhance the practical application of these tools, paving the way for more effective human–robot interaction systems aiming to improve productivity and safety in various industrial and collaborative settings.

Given the above information, readers should nevertheless take notice that the features and metrics mentioned and explained throughout this review are in no way meant to be taken as a fully exhaustive list of all the possible features and metrics of all sensors. Providing such an exhaustive list is beyond the scope of this review. For more information regarding features and metrics both mentioned and not mentioned in this review, please consult works such as [27,29].

## 5. Conclusions

The results of this systematic literature review highlight the potential of physiological metrics for monitoring mental workload. However, this study also identifies significant challenges in applying them to collaborative robotics, worsened by the lack of universal

consensus on its conceptualization and assessment. Diverse definitions of workload and a lack of standardization in monitoring methodologies further complicate the issue. While various metrics demonstrate the capability to measure mental workload, differences arise in the approach to gathering physiological variables. Some studies focus on a single metric, while others adopt a multimodal approach for a comprehensive understanding of mental workload in diverse collaborative scenarios.

The fact remains that no single metric currently exists that is exempt from limitations or being affected by outside noise, such as humidity, temperature, or a subject's medication. However, as technology progresses, one can expect the impact of these limitations to be decreased or eliminated outright. In the meanwhile, authors working with mental workload assessments have access to various tools to that end, allowing them to adapt to different use cases. For example, while current instruments to measure brain activity, such as the EEG and fNIRS, are prone to noise due to the subject's motion (e.g., walking around the room versus being stationary at a desk), wearable instruments that assess heart rate data while a subject is moving, such as smartwatches and smart rings, can be easily employed instead. While the heart rate measurement from these may be less reliable than other, more bulky instruments, they can nevertheless be good enough alternatives for situations where such instruments are not applicable or cumbersome, and the same can be said for other combinations of instruments as well.

As the field progresses, collaborative efforts among researchers and practitioners are crucial to defining robust methodologies, validating models across varied human–robot collaboration scenarios, and establishing guidelines for their implementation. This concerted approach will facilitate the development of machine learning models that not only enhance the accuracy of mental workload assessments but also ensure their applicability and reliability in real-world collaborative robotic environments. In the context of Industry 5.0, the utilization of various physiological variables as inputs for machine learning models holds promise for powering intelligent algorithms that can decode human movement intentions and adapt robot behavior proactively.

However, challenges persist in understanding the relationships among physiological variables and in effectively employing machine learning models to accurately assess mental workload. These challenges can be addressed by applying explainable artificial intelligence (XAI) to enable the interpretation and trust of the results generated by machine learning algorithms [81]. Further research is needed to evaluate the generalizability of the models across different tasks and to ensure high accuracy for industrial deployment. Addressing subject-specific factors, electrode placement variability, and changing environmental conditions in industrial settings will be essential to improve the practical implementation of cognitive workload classification systems.

In conclusion, while a universally validated method is not yet established, the foundation for future research lies in systematic description, ensuring the dependability of forthcoming models. Despite challenges, the integration of diverse methodologies holds significant potential in assessing mental workload and shaping the future of human–robot collaboration for enhanced safety, health, well-being, and productivity. This synthesis of findings emphasizes the need for continued exploration and refinement, paving the way for a more robust understanding and application of workload assessment methodologies in diverse contexts.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app15063317/s1, Table S1. Tasks description in developed collaborative scenarios.

**Author Contributions:** E.P.: Conceptualization, methodology, writing—original draft preparation; L.S.: methodology, writing—review and editing; E.S.: methodology, writing—review and editing; A.S., N.C. (Nuno Costa), and N.C. (Nélson Costa): writing—review and editing, supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work has received support from FCT—Fundação para a Ciência e Tecnologia within the PhD fellowship referenced as 2022.14626.BDANA. This work has been supported by FCT–Fundação para a Ciência e Tecnologia within the R&D Unit Project Scope UID/00319/Centro ALGORITMI (ALGORITMI/UM).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| ACC | Accelerometry |
| AOI | Area of Interest |
| BDM | Body Discomfort Map |
| BWS | Bedford Workload Scale |
| CP | Conference Paper |
| DFA | Detrended Fluctuation Analysis |
| DNN | Deep Neural Network |
| ECG | Electrocardiogram |
| EDA | Electrodermal Activity |
| EEG | Electroencephalogram |
| ERP | Error-Related Potential |
| fNIRS | Functional Near-Infrared Spectroscopy |
| GB | Gradient Boost |
| HF | High-Frequency |
| HRC | Human–Robot Collaboration |
| HRI | Human–Robot Interaction |
| HRVTi | Heart Rate Variability Time Index |
| HRV | Heart Rate Variability |
| IBI | Interbeat Interval |
| ISO | International Organization for Standardization |
| JP | Journal Paper |
| KNN | K-nearest Neighbors |
| LF | Low-Frequency |
| MWL | Mental Workload |
| NARS | Negative Attitude Towards Robots Survey |
| NASA-TLX | National Aeronautics and Space Administration Task Load Index |
| NN | Normal-to-Normal Intervals |
| OM | Ocular Measures |
| Pe | Error Positivity |
| PEN | Prediction Error Negativity |
| PNN50 | Proportion of Beats with a Successive R-R Interval Difference Exceeding 50 ms |
| PPG | Photoplethysmogram |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |

| | |
|---|---|
| PROSPERO | International Prospective Register of Systematic Reviews |
| RAS | Robot Anxiety Survey |
| RMSSD | Root Mean Square of the Successive R-R Interval Differences |
| SaEn | Sample Entropy |
| SAM | Self-Assessment Manikin |
| SCL | Skin Conductance Level |
| SCR | Skin Conductance Response |
| SDNN | Standard Deviation of all Successive Normal R-R Intervals |
| SIMKAP | Simultaneous Capacity Multitasking |
| SMI | Sympathetic Modulation Index |
| SPO2 | Peripheral Oxygen Saturation |
| SVI | Sympathovagal Balance Index |
| SVM | Support Vector Machines |
| TAM3 | Technology Acceptance Model |
| ULF | Ultra-Low-Frequency |
| VLF | Very-Low-Frequency |
| VMI | Vagal Modulation Index |
| XAI | Explainable Artificial Intelligence |

## References


1. Askarpour, M.; Mandrioli, D.; Rossi, M.; Vicentini, F. Formal Model of Human Erroneous Behavior for Safety Analysis in Collaborative Robotics. *Robot. Comput. Integr. Manuf.* **2019**, *57*, 465–476. [CrossRef]
2. Madonna, M.; Monica, L.; Anastasi, S.; Di Nardo, M. Evolution of Cognitive Demand in the Human–Machine Interaction Integrated with Industry 4.0 Technologies. In *WIT Transactions on the Built Environment*; WIT Press: Southampton, UK, 2019; Volume 189, pp. 13–19.
3. Kumar, S.; Savur, C.; Sahin, F. Survey of Human-Robot Collaboration in Industrial Settings: Awareness, Intelligence, and Compliance. *IEEE Trans. Syst. Man. Cybern. Syst.* **2021**, *51*, 280–297. [CrossRef]
4. Gervasi, R.; Capponi, M.; Mastrogiacomo, L.; Franceschini, F. Manual Assembly and Human–Robot Collaboration in Repetitive Assembly Processes: A Structured Comparison Based on Human-Centered Performances. *Int. J. Adv. Manuf. Technol.* **2023**, *126*, 1213–1231. [CrossRef]
5. Story, M.; Webb, P.; Fletcher, S.R.; Tang, G.; Jaksic, C.; Carberry, J. Do Speed and Proximity Affect Human-Robot Collaboration with an Industrial Robot Arm? *Int. J. Soc. Robot.* **2022**, *14*, 1087–1102. [CrossRef]
6. *ISO 10218-2:2011*; Robots and Robotic Devices—Safety Requirements for Industrial Robots—Part 2: Robot Systems and Integration. ISO: Geneva, Switzerland, 2011.
7. *ISO/TS15066:2016*; Robots and Robotic Devices—Collaborative Robots. ISO: Geneva, Switzerland, 2016.
8. Gervasi, R.; Capponi, M.; Mastrogiacomo, L.; Franceschini, F. Analyzing Psychophysical State and Cognitive Performance in Human-Robot Collaboration for Repetitive Assembly Processes. *Prod. Eng.* **2024**, *18*, 19–33. [CrossRef]
9. Longo, L.; Wickens, C.D.; Hancock, P.A.; Hancock, G.M. Human Mental Workload: A Survey and a Novel Inclusive Definition. *Front. Psychol.* **2022**, *13*, 883321. [CrossRef]
10. Pereira, E.; Costa, S.; Costa, N.; Arezes, P. Wellness in Cognitive Workload—A Conceptual Framework. In *Advances in Neuroergonomics and Cognitive Engineering*; Spring: Berlin/Heidelberg, Germany, 2019; pp. 353–364. ISBN 978-3-319-94865-2.
11. Coughlin, J.F.; Reimer, B.; Mehler, B. *Driver Wellness, Safety & the Development of an Awarecar*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2009; pp. 1–15.
12. Gervasi, R.; Capponi, M.; Antonelli, D.; Mastrogiacomo, L.; Franceschini, F. A Human-Centered Perspective in Repetitive Assembly Processes: Preliminary Investigation of Cognitive Support of Collaborative Robots. *Procedia Comput. Sci.* **2024**, *232*, 2249–2258. [CrossRef]
13. McKendrick, R.; Feest, B.; Harwood, A.; Falcone, B. Theories and Methods for Labeling Cognitive Workload: Classification and Transfer Learning. *Front. Hum. Neurosci.* **2019**, *13*, 295. [CrossRef]
14. *ISO 10075-3:2004*; Ergonomic Principles Related to Mental Workload Part 3: Principles and Requirements Concerning Methods for Measuring and Assessing Mental Workload. ISO: Geneva, Switzerland, 2004.
15. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*; Hancock, P.A., Meshkati, N., Eds.; Advances in Psychology; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183.

16. Filho, P.C.A.; da Silva, L.; Pombeiro, A.; Costa, N.; Carneiro, P.; Arezes, P. Assessing Mental Workload in Industrial Environments: A Review of Applied Studies. *Stud. Syst. Decis. Control* **2024**, *492*, 677–689. [CrossRef]
17. Costa, N.; Costa, S.; Pereira, E.; Arezes, P. Workload Measures—Recent Trends in the Driving Context. In *Studies in Systems, Decision and Control*; Spring: Berlin/Heidelberg, Germany, 2019; pp. 419–430. ISBN 978-3-658-25495-7.
18. Muñoz, J.E.; Pereira, F.; Karapanos, E. Workload Management through Glanceable Feedback: The Role of Heart Rate Variability. In Proceedings of the 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services, Healthcom 2016, Munich, Germany, 14–17 September 2016.
19. Landi, C.T.; Villani, V.; Ferraguti, F.; Sabattini, L.; Secchi, C.; Fantuzzi, C. Relieving Operators' Workload: Towards Affective Robotics in Industrial Scenarios. *Mechatronics* **2018**, *54*, 144–154. [CrossRef]
20. Lagomarsino, M.; Lorenzini, M.; Balatti, P.; De Momi, E.; Ajoudani, A. Pick the Right Co-Worker: Online Assessment of Cognitive Ergonomics in Human-Robot Collaborative Assembly. *IEEE Trans. Cogn. Dev. Syst.* **2022**, *15*, 1928–1937. [CrossRef]
21. Mao, X.; Li, M.; Li, W.; Niu, L.; Xian, B.; Zeng, M.; Chen, G. Progress in EEG-Based Brain Robot Interaction Systems. *Comput. Intell. Neurosci.* **2017**, *2017*, 1742862. [CrossRef]
22. Panchetti, T.; Pietrantoni, L.; Puzzo, G.; Gualtieri, L.; Fraboni, F. Assessing the Relationship between Cognitive Workload, Workstation Design, User Acceptance and Trust in Collaborative Robots. *Appl. Sci.* **2023**, *13*, 1720. [CrossRef]
23. Nenna, F.; Orso, V.; Zanardi, D.; Gamberini, L. The Virtualization of Human–Robot Interactions: A User-Centric Workload Assessment. *Virtual Real.* **2023**, *27*, 553–571. [CrossRef]
24. Rajavenkatanarayanan, A.; Nambiappan, H.R.; Kyrarini, M.; Makedon, F. Towards a Real-Time Cognitive Load Assessment System for Industrial Human-Robot Cooperation. In Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; pp. 698–705.
25. Buerkle, A.; Matharu, H.; Al-Yacoub, A.; Lohse, N.; Bamber, T.; Ferreira, P. An Adaptive Human Sensor Framework for Human–Robot Collaboration. *Int. J. Adv. Manuf. Technol.* **2022**, *119*, 1233–1248. [CrossRef]
26. Bettoni, A.; Montini, E.; Righi, M.; Villani, V.; Tsvetanov, R.; Borgia, S.; Secchi, C.; Carpanzano, E. Mutualistic and Adaptive Human-Machine Collaboration Based on Machine Learning in an Injection Moulding Manufacturing Line. *Procedia CIRP* **2020**, *93*, 395–400.
27. Tao, D.; Tan, H.; Wang, H.; Zhang, X.; Qu, X.; Zhang, T. A Systematic Review of Physiological Measures of Mental Workload. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2716. [CrossRef]
28. Lohani, M.; Payne, B.R.; Strayer, D.L. A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving. *Front. Hum. Neurosci.* **2019**, *13*, 57. [CrossRef]
29. Charles, R.L.; Nixon, J. Measuring Mental Workload Using Physiological Measures: A Systematic Review. *Appl. Ergon.* **2019**, *74*, 221–232. [CrossRef]
30. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71.
31. McHugh, M.L. Interrater Reliability: The Kappa Statistic. *Biochem. Med.* **2012**, *22*, 276–282. [CrossRef]
32. Liu, Y.; Habibnezhad, M.; Jebelli, H. Brainwave-Driven Human-Robot Collaboration in Construction. *Autom. Constr.* **2021**, *124*, 103556. [CrossRef]
33. Liu, Y.; Habibnezhad, M.; Jebelli, H.; Monga, V. Worker-in-the-Loop Cyber-Physical System for Safe Human-Robot Collaboration in Construction. In Proceedings of the Computing in Civil Engineering 2021—Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021, Orlando, FL, USA, 12–14 September 2021; American Society of Civil Engineers (ASCE): Reston, VA, USA, 2021; pp. 1075–1083.
34. Lin, C.J.; Lukodono, R.P. Classification of Mental Workload in Human-Robot Collaboration Using Machine Learning Based on Physiological Feedback. *J. Manuf. Syst.* **2022**, *65*, 673–685. [CrossRef]
35. Liu, Y.; Jebelli, H. Worker-Aware Robotic Motion Planner in Construction for Improved Psychological Well- Being during Worker-Robot Interaction. In *Construction Research Congress 2022: Computer Applications, Automation, and Data Analytics—Selected Papers from Construction Research Congress 2022*; American Society of Civil Engineers (ASCE): Reston, VA, USA, 2022; Volume 2-B, pp. 205–214.
36. Savković, M.; Caiazzo, C.; Djapan, M.; Vukićević, A.M.; Pušica, M.; Mačužić, I. Development of Modular and Adaptive Laboratory Set-Up for Neuroergonomic and Human-Robot Interaction Research. *Front. Neurorobot.* **2022**, *16*, 863637. [CrossRef]
37. Afzal, M.A.; Gu, Z.; Afzal, B.; Bukhari, S.U. Cognitive Workload Classification in Industry 5.0 Applications: Electroencephalography-Based Bi-Directional Gated Network Approach. *Electronics* **2023**, *12*, 4008. [CrossRef]
38. Caiazzo, C.; Savkovic, M.; Pusica, M.; Milojevic, D.; Leva, M.C.; Djapan, M. Development of a Neuroergonomic Assessment for the Evaluation of Mental Workload in an Industrial Human–Robot Interaction Assembly Task: A Comparative Case Study. *Machines* **2023**, *11*, 995. [CrossRef]

39. Liu, Y.; Ojha, A.; Shayesteh, S.; Jebelli, H.; Lee, S. Human-Centric Robotic Manipulation in Construction: Generative Adversarial Networks Based Physiological Computing Mechanism to Enable Robots to Perceive Workers' Cognitive Load. *Can. J. Civ. Eng.* **2023**, *50*, 224–238. [CrossRef]
40. Pluchino, P.; Pernice, G.F.A.; Nenna, F.; Mingardi, M.; Bettelli, A.; Bacchin, D.; Spagnolli, A.; Jacucci, G.; Ragazzon, A.; Miglioranzi, L.; et al. Advanced Workstations and Collaborative Robots: Exploiting Eye-Tracking and Cardiac Activity Indices to Unveil Senior Workers' Mental Workload in Assembly Tasks. *Front. Robot. AI* **2023**, *10*, 1275572. [CrossRef]
41. Stanković, E.; Kljajić, J.; Šumarać, J.; Radmilović, M. Human Motion Behavior Evaluation: The Possibility of Improving Human-Robot Collaboration. In Proceedings of the 10th International Conference on Electrical, Electronic and Computing Engineering, IcETRAN 2023, East Sarajevo, RS, Bosnia and Herzegovina, 5–8 June 2023; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2023.
42. Zakeri, Z.; Arif, A.; Omurtag, A.; Breedon, P.; Khalid, A. Multimodal Assessment of Cognitive Workload Using Neural, Subjective and Behavioural Measures in Smart Factory Settings. *Sensors* **2023**, *23*, 8926. [CrossRef]
43. Capponi, M.; Gervasi, R.; Mastrogiacomo, L.; Franceschini, F. Assembly Complexity and Physiological Response in Human-Robot Collaboration: Insights from a Preliminary Experimental Analysis. *Robot Comput. Integr. Manuf.* **2024**, *89*, 102789. [CrossRef]
44. Chen, S.-Y.; Chen, J.-H.; Yang, B.-S. Investigating Human Mental Workload During Human Robot Interaction. In Proceedings of the 2024 19th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT), Taipei, Taiwan, 22–25 October 2024; IEEE: New York, NY, USA, 2024; pp. 311–314.
45. Gervasi, R.; Capponi, M.; Mastrogiacomo, L.; Franceschini, F. Eye-Tracking Support for Analyzing Human Factors in Human-Robot Collaboration during Repetitive Long-Duration Assembly Processes. *Prod. Eng.* **2024**, *19*, 47–64. [CrossRef]
46. John, A.R.; Singh, A.K.; Gramann, K.; Liu, D.; Lin, C.T. Prediction of Cognitive Conflict during Unexpected Robot Behavior under Different Mental Workload Conditions in a Physical Human-Robot Collaboration. *J. Neural Eng.* **2024**, *21*, 026010. [CrossRef] [PubMed]
47. Kneževic, N.; Savić, A.; Gordić, Z.; Ajoudani, A.; Jovanović, K. Toward Industry 5.0: A Neuroergonomic Workstation for a Human-Centered, Collaborative Robot-Supported Manual Assembly Process. *IEEE Robot. Autom. Mag.* **2024**, 2–13. [CrossRef]
48. Korivand, S.; Galvani, G.; Ajoudani, A.; Gong, J.; Jalili, N. Optimizing Human–Robot Teaming Performance through Q-Learning-Based Task Load Adjustment and Physiological Data Analysis. *Sensors* **2024**, *24*, 2817. [CrossRef] [PubMed]
49. Packy, A.L.; Jayahankar, J.; Teymourlouei, A.; Stone, J.; Oh, H.; Katz, G.E.; Reggia, J.A.; Gentili, R.J. Neurocognitive Assessment under Various Human-Robot Teaming Environments. In Proceedings of the 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 15–19 July 2024; IEEE: New York, NY, USA, 2024; pp. 1–4.
50. Roesler, E.; Meerwein, J.; Krueger, J.; Onnasch, L. Beyond the Default: The Effects of Adaptable Robot Speed in Industrial Human-Robot Interaction. In Proceedings of the Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, Boulder, CO, USA, 11–15 March 2024; ACM: New York, NY, USA, 2024; pp. 896–900.
51. Segura, P.; Lobato-Calleros, O.; Soria-Arguello, I.; Hernández-Martínez, E.G. Work Roles in Human–Robot Collaborative Systems: Effects on Cognitive Ergonomics for the Manufacturing Industry. *Appl. Sci.* **2025**, *15*, 744. [CrossRef]
52. Yerebakan, M.O.; Gu, Y.; Gross, J.; Hu, B. Evaluation of Biomechanical and Mental Workload During Human–Robot Collaborative Pollination Task. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2025**, *67*, 100–114. [CrossRef]
53. van Eck, N.J.; Waltman, L. *VOSviewer Manual Version 1.6.16*; Univeristeit Leiden: Leiden, The Netherlands, 2020.
54. Iarlori, S.; Perpetuini, D.; Tritto, M.; Cardone, D.; Tiberio, A.; Chinthakindi, M.; Filippini, C.; Cavanini, L.; Freddi, A.; Ferracuti, F.; et al. An Overview of Approaches and Methods for the Cognitive Workload Estimation in Human–Machine Interaction Scenarios through Wearables Sensors. *BioMedInformatics* **2024**, *4*, 1155–1173. [CrossRef]
55. Jebelli, H.; Hwang, S.; Lee, S. EEG Signal-Processing Framework to Obtain High-Quality Brain Waves from an Off-the-Shelf Wearable EEG Device. *J. Comput. Civ. Eng.* **2018**, *32*, 04017070. [CrossRef]
56. Yang, J.; Barragan, J.A.; Farrow, J.M.; Sundaram, C.P.; Wachs, J.P.; Yu, D. An Adaptive Human-Robotic Interaction Architecture for Augmenting Surgery Performance Using Real-Time Workload Sensing—Demonstration of a Semi-Autonomous Suction Tool. *Hum. Factors* **2022**, *66*, 1081–1102. [CrossRef]
57. Betts, K.; Reddy, P.; Galoyan, T.; Delaney, B.; McEachron, D.L.; Izzetoglu, K.; Shewokis, P.A. An Examination of the Effects of Virtual Reality Training on Spatial Visualization and Transfer of Learning. *Brain Sci.* **2023**, *13*, 890. [CrossRef]
58. Strait, M.; Scheutz, M. What We Can and Cannot (yet) Do with Functional near Infrared Spectroscopy. *Front. Neurosci.* **2014**, *8*, 117. [CrossRef]
59. Cao, J.; Garro, E.M.; Zhao, Y. EEG/FNIRS Based Workload Classification Using Functional Brain Connectivity and Machine Learning. *Sensors* **2022**, *22*, 7623. [CrossRef]
60. Park, J.H. Mental Workload Classification Using Convolutional Neural Networks Based on FNIRS-Derived Prefrontal Activity. *BMC Neurol.* **2023**, *23*, 1–8. [CrossRef] [PubMed]

61. Aghajani, H.; Garbey, M.; Omurtag, A. Measuring Mental Workload with EEG+fNIRS. *Front. Hum. Neurosci.* **2017**, *11*, 359. [CrossRef] [PubMed]
62. Shao, S.; Wang, T.; Wang, Y.; Su, Y.; Song, C.; Yao, C. Research of Hrv as a Measure of Mental Workload in Human and Dual-Arm Robot Interaction. *Electronics* **2020**, *9*, 2174. [CrossRef]
63. Shaffer, F.; Ginsberg, J.P. An Overview of Heart Rate Variability Metrics and Norms. *Front. Public Health* **2017**, *5*, 258. [CrossRef]
64. Memar, A.H.; Esfahani, E.T. Human Performance in a Mixed Human-Robot Team: Design of a Collaborative Framework. In Proceedings of the ASME Design Engineering Technical Conference, Charlotte, NC, USA, 21–24 August 2016; American Society of Mechanical Engineers (ASME): New York, NY, USA, 2016; Volume 1B-2016.
65. Ghaderyan, P.; Abbasi, A. An Efficient Automatic Workload Estimation Method Based on Electrodermal Activity Using Pattern Classifier Combinations. *Int. J. Psychophysiol.* **2016**, *110*, 91–101. [CrossRef]
66. Saha, S.; Jindal, K.; Shakti, D.; Tewary, S.; Sardana, V. Chirplet Transform-Based Machine-Learning Approach towards Classification of Cognitive State Change Using Galvanic Skin Response and Photoplethysmography Signals. *Expert Syst.* **2022**, *39*, e12958. [CrossRef]
67. Posada-Quintero, H.F.; Chon, K.H. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. *Sensors* **2020**, *20*, 479. [CrossRef]
68. Han, S.Y.; Kwak, N.S.; Oh, T.; Lee, S.W. Classification of Pilots' Mental States Using a Multimodal Deep Learning Network. *Biocybern. Biomed. Eng.* **2020**, *40*, 324–336. [CrossRef]
69. Anusha, A.S.; Jose, J.; Preejith, S.P.; Jayaraj, J.; Mohanasankar, S. Physiological Signal Based Work Stress Detection Using Unobtrusive Sensors. *Biomed. Phys. Eng. Express* **2018**, *4*, 065001. [CrossRef]
70. Jimenez-Molina, A.; Retamal, C.; Lira, H. Using Psychophysiological Sensors to Assess Mental Workload during Web Browsing. *Sensors* **2018**, *18*, 458. [CrossRef]
71. DellrAgnola, F.; Jao, P.-K.; Arza, A.; Chavarriaga, R.; Millan, J.d.R.; Floreano, D.; Atienza, D. Machine-Learning Based Monitoring of Cognitive Workload in Rescue Missions with Drones. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4751–4762. [CrossRef] [PubMed]
72. Androutsou, T.; Angelopoulos, S.; Hristoforou, E.; Matsopoulos, G.K.; Koutsouris, D.D. A Multisensor System Embedded in a Computer Mouse for Occupational Stress Detection. *Biosensors* **2023**, *13*, 10. [CrossRef] [PubMed]
73. Ding, Y.; Cao, Y.; Wang, Y. Physiological Indicators of Mental Workload in Visual Display Terminal Work. In *Advances in Physical Ergonomics and Human Factors*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; Volume 967, ISBN 9783030201418.
74. De Waard, D. The Measurement of Drivers's Mental Workload. Ph.D. Thesis, University of Groningen, Traffic Research Centre, Groningen, The Netherlands, 1996.
75. Kim, H.; Miyakoshi, M.; Kim, Y.; Stapornchaisit, S.; Yoshimura, N.; Koike, Y. Electroencephalography Reflects User Satisfaction in Controlling Robot Hand through Electromyographic Signals. *Sensors* **2023**, *23*, 277. [CrossRef] [PubMed]
76. Hopko, S.K.; Khurana, R.; Mehta, R.K.; Pagilla, P.R. Effect of Cognitive Fatigue, Operator Sex, and Robot Assistance on Task Performance Metrics, Workload, and Situation Awareness in Human-Robot Collaboration. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3049–3056. [CrossRef]
77. Shayesteh, S.; Ojha, A.; Liu, Y.; Jebelli, H. Human-Robot Teaming in Construction: Evaluative Safety Training through the Integration of Immersive Technologies and Wearable Physiological Sensing. *Saf. Sci.* **2023**, *159*, 106019. [CrossRef]
78. Gervasi, R.; Aliev, K.; Mastrogiacomo, L.; Franceschini, F. User Experience and Physiological Response in Human-Robot Collaboration: A Preliminary Investigation. *J. Intell. Robot. Syst. Theory Appl.* **2022**, *106*, 36. [CrossRef]
79. Landmann, E. I Can See How You Feel—Methodological Considerations and Handling of Noldus's FaceReader Software for Emotion Measurement. *Technol. Forecast Soc. Change* **2023**, *197*, 122889. [CrossRef]
80. Zhou, Y.; Xu, Z.; Niu, Y.; Wang, P.; Wen, X.; Wu, X.; Zhang, D. Cross-Task Cognitive Workload Recognition Based on EEG and Domain Adaptation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 50–60. [CrossRef]
81. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence. *Inf. Fusion* **2023**, *99*, 101805. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.