DEMOCRITUS UNIVERSITY OF THRACE
FACULTY OF HEALTH SCIENCES
SCHOOL OF MEDICINE

ATHENA
RESEARCH AND INNOVATION CENTER IN INFORMATION COMMUNICATION AND KNOWLEDGE TECHNOLOGIES

Digital Biomarkers of Stress and Anxiety: Assessing the Quality and Diversity of Wearable Sensor Datasets

MSc THESIS

MSc in

BIOMEDICAL INFORMATICS

Georgia Kalitsi

Supervisor:
Eleni Kaldoudi
Professor of Medical Physics and Medical Informatics
School of Medicine, Democritus University of Thrace, Greece

ALEXANDROUPOLI, GREECE

MARCH 2026

MSc Thesis

Submitted for the Degree of Master of Science in Biomedical Informatics,

Jointly Organized by the

School of Medicine, Democritus University of Thrace, Greece and

ATHENA Research and Innovation Center in Information Communication and Knowledge Technologies, Greece

Digital Biomarkers of Stress and Anxiety: Assessing the Quality and Diversity of Wearable Sensor Datasets

by

Georgia Kalitsi

Supervised by

Eleni Kaldoudi

Professor of Medical Physics and Medical Informatics
School of Medicine
Democritus University of Thrace, Greece

Approved on 26 March 2026 by the Examination Committee:

Eleni Kaldoudi	**Stylianos Didaskalou**	**Aliki Fiska**
Professor School of Medicine, Democritus University of Thrace, Greece	Associate Researcher ATHENA Research and Innovation, Greece	Professor School of Medicine, Democritus University of Thrace, Greece

Μεταπτυχιακή Διπλωματική Εργασία

Υποβλήθηκε για το Πτυχίο Μεταπτυχιακών Σπουδών στην Βιοϊατρική Πληροφορική

που Συνδιοργανώνεται από το

Τμήμα Ιατρικής, Δημοκρίτειο Πανεπιστήμιο Θράκης και

ΑΘΗΝΑ Ερευνητικό Κέντρο Καινοτομίας στις Τεχνολογίες της Πληροφορίας, των Επικοινωνιών και της Γνώσης

Ψηφιακοί Βιοδείκτες για Αγχώδεις Διαταραχές: Αξιολόγηση Ποιότητας και Ετερογένειας Δεδομένων από Φορετούς Αισθητήρες

της

Γεωργίας Καλίτση

Επιβλέπουσα

Ελένη Καλδούδη

Καθηγήτρια Ιατρικής Φυσικής – Ιατρικής Πληροφορικής
Τμήμα Ιατρικής, Δημοκρίτειο Πανεπιστήμιο Θράκης

Εγκρίθηκε στις 26 Μαρτίου 2026 από την Εξεταστική Επιτροπή:

Ελένη Καλδούδη	**Στυλιανός Διδασκάλου**	**Αλίκη Φίσκα**
Καθηγήτρια Τμήμα Ιατρικής, Δημοκρίτειο Πανεπιστήμιο Θράκης	Συνεργαζόμενος Ερευνητής Ερευνητικό Κέντρο ΑΘΗΝΑ	Καθηγήτρια Τμήμα Ιατρικής, Δημοκρίτειο Πανεπιστήμιο Θράκης

Kalitsi G, Digital Biomarkers of Stress and Anxiety: Assessing the Quality and Diversity of Wearable Sensor Datasets, MSc Thesis, Master of Science in Biomedical Informatics, School of Medicine, Democritus University of Thrace, and ATHENA Research Center, Alexandroupoli, Greece, March 2026. https://doi.org/10.5281/zenodo.19006535

The author declares that this thesis is his/her own work and to the best of his/her knowledge it:

- Does not breach copyright or other intellectual property rights of a third party.
- Does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
- Does not contain material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.
- Does not contain substantial portions of third-party copyright material, including but not limited to charts, diagrams, graphs, photographs or maps, or in instances where it does the author has obtained permission to use such material and allow it to be made accessible worldwide via the Internet as part of this thesis.

During the preparation of this thesis Gemini 3 Flash, March 2026 was used to assist with language refinement, content structuring, and editing support. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the final thesis.

The contents of this thesis are solely the responsibility of the author. Approval of this thesis does not necessarily represent the official views of the Examination Committee, the School of Medicine, Democritus University of Thrace, or ATHENA Research and Innovation Center in Information Communication and Knowledge Technologies.

Καλίτση Γ., Ψηφιακοί Βιοδείκτες για Αγχώδεις Διαταραχές: Αξιολόγηση Ποιότητας και Ετερογένειας Δεδομένων από Φορετούς Αισθητήρες, Μεταπτυχιακή Διπλωματική Εργασία, ΔΠΜΣ Βιοϊατρική Πληροφορική, Τμήμα Ιατρικής, ΔΠΘ και ΑΘΗΝΑ ΕΚ, Αλεξανδρούπολη, Μάρτιος, 2026. https://doi.org/10.5281/zenodo.19006535

Κατά τη συγγραφή της παρούσας διατριβής χρησιμοποιήθηκε το εργαλείο τεχνητής νοημοσύνης Gemini 3 Flash, March 2026 ως βοήθημα για βελτίωση της σύνταξης και της δομής του κειμένου και υποστήριξη της επιμέλειας. Μετά τη χρήση του εργαλείου, ο/η συγγραφέας επιμελήθηκε το περιεχόμενο και φέρει την πλήρη ευθύνη για το περιεχόμενο της τελικής εργασίας.

Τα περιεχόμενα αυτής της διατριβής είναι αποκλειστική ευθύνη του/της συγγραφέα. Η έγκριση της διατριβής Θράκης δεν δηλώνει απαραιτήτως την αποδοχή των απόψεων του συγγραφέα από την Εξεταστική Επιτροπή, το Τμήμα Ιατρικής του Δημοκριτείου Πανεπιστημίου, ή το ΑΘΗΝΑ Ερευνητικό Κέντρο Καινοτομίας στις Τεχνολογίες της Πληροφορίας, των Επικοινωνιών και της Γνώσης.

Digital Biomarkers of Stress and Anxiety: Assessing the Quality and Diversity of Wearable Sensor Datasets

Georgia Kalitsi

SUMMARY

This thesis explores the rapidly evolving field of digital biomarkers, focusing on the use of wearable sensors for the detection and monitoring of stress and anxiety. As the global burden of mental health conditions increases, wearable technology offers a transformative opportunity for the continuous, passive, and objective monitoring of physiological states in real-world settings. However, despite the proliferation of research in this domain, the field remains characterized by extreme heterogeneity. Datasets generated by various research groups often utilize different hardware, sampling frequencies, and stress-induction methodologies, creating a harmonization gap that prevents the development of generalizable, sensor-agnostic artificial intelligence models.

The theoretical foundation of this work is built upon the evolution of stress research, moving from traditional clinical snapshots and biochemical markers toward modern autonomic nervous system monitoring. The research identifies the core signals necessary for robust digital biomarkers via a systematic search, following the PRISMA 2020 statement, targeting peer-reviewed studies published over the last decade. This search, performed via PubMed, identified 63 unique datasets which were then meticulously mapped based on their technical specifications, cohort demographics, and accessibility.

Results indicate that the field of digital stress monitoring has transitioned from exploratory studies to an era of intensive data collection driven by the maturation of wearable technology, yet it remains hindered by significant data accessibility barriers and a persistent reliance on controlled laboratory environments over real-world settings. While physiological sensing is anchored by modalities like Electrodermal Activity (EDA), heart rate, and accelerometry, there is a clear shift toward higher-dimensional datasets and a strategic preference for more robust, motion-resistant signals in "in-the-wild" applications to ensure ecological validity. Despite the standardization of laboratory stressors and the frequent use of self-reported metrics for annotation, the extreme heterogeneity in hardware and methodologies has created a harmonization gap that prevents the fusion of these disparate data sources.

To address this, the thesis proposes a 7-Tier Harmonization Roadmap as a structured, hierarchical framework to integrate heterogeneous data and bridge the gap between technical interoperability and physiological compatibility across diverse sensing ecosystems. While the study is limited by its primary focus on PubMed-indexed literature and the inherent inconsistencies in "ground truth" labeling across different studies, it provides a foundational strategy for the scientific community. The work signifies a shift from sensor-specific silos toward a unified data ecosystem. Ultimately, this thesis argues that the future of reliable digital stress biomarkers depends not on the creation of more isolated datasets, but on the systematic harmonization of existing resources to build truly generalizable and clinically relevant diagnostic tools.

Ψηφιακοί Βιοδείκτες για Αγχώδεις Διαταραχές: Αξιολόγηση Ποιότητας και Ετερογένειας Δεδομένων από Φορετούς Αισθητήρες

Γεωργία Καλίτση

ΠΕΡΙΛΗΨΗ

Η παρούσα διατριβή διερευνά το ταχέως εξελισσόμενο πεδίο των ψηφιακών βιοδεικτών, εστιάζοντας στη χρήση φορετών αισθητήρων για την ανίχνευση και παρακολούθηση του στρες και του άγχους. Καθώς η παγκόσμια επιβάρυνση από παθήσεις της ψυχικής υγείας αυξάνεται, η τεχνολογία των φορετών συσκευών προσφέρει μια καινοτόμο ευκαιρία για συνεχή, παθητική και αντικειμενική παρακολούθηση της φυσιολογικής κατάστασης του ατόμου σε συνθήκες καθημερινής δραστηριότητας. Ωστόσο, το πεδίο χαρακτηρίζεται από έντονη ετερογένεια. Τα σύνολα δεδομένων που παράγονται από διαφορετικές ερευνητικές ομάδες συχνά χρησιμοποιούν διαφορετικό υλικό (hardware), συχνότητες δειγματοληψίας και μεθοδολογίες πρόκλησης στρες, δημιουργώντας ένα «χάσμα εναρμόνισης» που εμποδίζει την ανάπτυξη γενικευμένων μοντέλων τεχνητής νοημοσύνης, ανεξάρτητων από τον τύπο του αισθητήρα.

Η παρούσα έρευνα επιχειρεί την αποτύπωση της σημερινής κατάστασης σε διαθέσιμα σύνολα δεδομένων που αφορούν σε ψηφιακούς βιοδείκτες για αγχώδεις διαταραχές μέσα από μια συστηματική ανασκόπηση της βιβλιογραφίας. Η συστηματική αναζήτηση ακολούθησε το πρωτόκολλο PRISMA 2020 και στόχευσε σε επιστημονικές μελέτες που δημοσιεύτηκαν την τελευταία δεκαετία. Η αναζήτηση, η οποία πραγματοποιήθηκε στη βάση δεδομένων PubMed, ανέδειξε 63 μοναδικά σύνολα δεδομένων, τα οποία στη συνέχεια χαρτογραφήθηκαν σχολαστικά με βάση τις τεχνικές προδιαγραφές, τα δημογραφικά στοιχεία των συμμετεχόντων και τη δυνατότητα πρόσβασης σε αυτά.

Τα αποτελέσματα δείχνουν ότι το πεδίο της ψηφιακής παρακολούθησης του στρες έχει μεταβεί από το στάδιο των σποραδικών διερευνητικών μελετών σε μια εποχή εντατικής συλλογής δεδομένων, ωθούμενο από την ωρίμανση της τεχνολογίας των φορετών συσκευών. Παρόλα αυτά, η πρόοδος παρεμποδίζεται από σημαντικούς περιορισμούς στην πρόσβαση στα δεδομένα και από μια συνεχή εξάρτηση από ελεγχόμενα εργαστηριακά περιβάλλοντα έναντι των πραγματικών συνθηκών διαβίωσης. Ενώ η καταγραφή φυσιολογικών παραμέτρων βασίζεται σε μεθόδους όπως η ηλεκτροδερμική δραστηριότητα (EDA), ο καρδιακός ρυθμός και η επιταχυνσιομετρία, παρατηρείται μια σαφής στροφή προς πολυδιάστατα σύνολα δεδομένων και μια στρατηγική προτίμηση σε σήματα πιο ανθεκτικά σε παρεμβολές από την κίνηση για εφαρμογές σε πραγματικές συνθήκες. Παρά την τυποποίηση των εργαστηριακών παραγόντων άγχους και τη συχνή χρήση υποκειμενικών αναφορών για τον χαρακτηρισμό του στρες, η ετερογένεια στον εξοπλισμό και τις μεθοδολογίες έχει δημιουργήσει ένα κενό εναρμόνισης που εμποδίζει τη συγχώνευση αυτών των ανόμοιων πηγών δεδομένων.

Για την αντιμετώπιση αυτού του ζητήματος, η διατριβή προτείνει μια προσέγγιση εναρμόνισης. Πρόκειται για ένα δομημένο, ιεραρχικό πλαίσιο για την ενσωμάτωση ετερογενών δεδομένων, το οποίο γεφυρώνει το κενό μεταξύ τεχνικής διαλειτουργικότητας και φυσιολογικής συμβατότητας σε διαφορετικά οικοσυστήματα αισθητήρων. Αν και η μελέτη περιορίζεται από την εστίασή της στην βιβλιογραφία της βάσης PubMed και από τις εγγενείς ασυνέπειες στον τρόπο χαρακτηρισμού των δεδομένων (ground truth labeling), παρέχει μια

θεμελιώδη στρατηγική για την επιστημονική κοινότητα. Η εργασία σηματοδοτεί τη μετάβαση από στεγανοποιημένες προσεγγίσεις που βασίζονται σε συγκεκριμένους αισθητήρες προς ένα ενιαίο οικοσύστημα δεδομένων. Τελικά, η διατριβή υποστηρίζει ότι το μέλλον των αξιόπιστων ψηφιακών βιοδεικτών στρες δεν εξαρτάται από τη δημιουργία περισσότερων μεμονωμένων συνόλων δεδομένων, αλλά από τη συστηματική εναρμόνιση των υπαρχόντων πόρων για την οικοδόμηση πραγματικά γενικευμένων και κλινικά αξιοποιήσιμων διαγνωστικών εργαλείων.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my academic supervisor, Professor Eleni Kaldoudi, for her invaluable guidance, continuous support and for giving me the opportunity to explore the field of biomedical informatics. Her expertise and her insightful feedback throughout the development of this research have been fundamental to the completion of this thesis.

I am also sincerely grateful to the members of my examination committee, Dr. Stylianos Didaskalou and Professor Aliki Fiska, for their time, interest and the constructive evaluation of this work.

I would like to thank the School of Medicine of the Democritus University of Thrace and the ATHENA Research and Innovation Center for providing an environment for my postgraduate studies. The interdisciplinary character of the MSc program in Biomedical Informatics was a truly transformative experience for my academic and professional development.

On a more personal note, I would like to thank my family, who have always been a source of inspiration and motivation for me. My family's love, patience, and trust in me have been the biggest strengths of mine. This achievement would have never been possible without the support of my family.

Georgia Kalitsi, Alexandroupoli, March 2026

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
3D	Three dimensional
ACC	Acceleration
ACTH	Adrenocorticotropic Hormone
AI	Artificial Intelligence
ANS	Autonomic Nervous System
BIS/BAS	Behavioral Inhibition System and Behavioral Activation System Scales
BVP	Blood Volume Pulse
CRH	Corticotropin-releasing Hormone
CSF	Chalder Fatigue Scale
DASS	Depression, Anxiety, and Stress Scale
DHEA	Dehydroepiandrosterone
DSI	Daily Stress Inventory
DSM	Diagnostic and Statistical Manual of Mental Disorders
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalography
EMA	Ecological Momentary Assessment
EMG	Electromyography
GAD	Generalized Anxiety Disorder Screening Tool
GSR	Galvanic Skin Response
HAM-A	Hamilton Anxiety Rating Scale
HF-HRV	High Frequency Heart Rate Variability
HPA	Hypothalamic–pituitary–adrenal axis
HR	Heart Rate
HRV	Heart Rate Variability
IAPS	International Affective Picture System
IBI	Interbeat intervals
ICD	International Classification of Diseases
LEDS	Life Events and Difficulties Scedule
LSAS-SR	Liebowitz Social Anxiety Scale - Self-Report

Abbreviation	Description
MeSH	Medial Subject Headings
NASA-TLX	NASA Task Load Index
NML	National Library of Medicine
PANAS	Positive and Negative Affect Schedule
PHQ	Patient Health Questionnaire
PMID	PubMed Unique Identifier
PPG	Photoplethysmography
PR	PR interval in ECG: the interval between the P wave and the QRS complex
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PRV	Pulse Rate Variability
PSQI	Pittsburgh Sleep Quality Index
PSS	Perceived Stress Scale
PTSD	Post-Traumatic Stress Disorder
QRS	QRS complex in ECG: the Q wave, the R wave, and the S wave
QTc	Corrected QT interval in ECG:, where QT denotes the total ventricular cycle
RESP	Respiration
RIP	Respiratory Inductance Plethysmography
RMSSD	Root Mean Square of Successive Differences
RPPG	Remote Photoplethysmography
RSME	Rating Scale Mental Effort
SAM	Self-Assessment Manikins
S-Anxiety	State Anxiety
SAS	Sympathetic Adrenal System
SCWT	Stroop Color and Word Test
SFS	Social Functioning Scale
SNS	Sympathetic Nervous System
SPSQ	Social Phobia Screening Questionnaire
SSSQ	Short Stress State Questionnaire
STAI	State-Trait Anxiety Inventory
STRAIN	Stress and Adversity Inventory
SUDS	Subjective Units of Distress Scale
T-Anxiety	Trait Anxiety

Abbreviation	Description
TEMP	Temperature
TICS	Trier Inventory for the Assessment of Chronic Stress
TSST	Trier Social Stress Test
USA	United States of America
VAD	Valence-Arousal-Dominance Model
VAS	Visual Analogue Scale
WAI	Wearable Artificial Intelligence
WHO	World Health Organization

Chapter 1| Introduction

The global burden of mental health conditions has reached unprecedented levels, with stress and anxiety serving as primary contributors to long-term morbidity and reduced quality of life. Traditional clinical assessment of these conditions has historically relied on episodic, snapshot evaluations, such as clinical interviews and retrospective self-report questionnaires. While these methods are scientifically validated, they are often limited by recall bias and the "white coat" effect, where the clinical environment itself acts as an extraneous stressor. Furthermore, these methods fail to capture the dynamic, fluctuating nature of physiological stress responses as they occur in a participant's natural environment.

The rapid proliferation of wearable technology devices has led in a massive generation of physiological data, primarily aimed at stress and anxiety detection. Nevertheless, the current research environment is characterized by extreme heterogeneity. Existing datasets differ considerably in terms of hardware characteristics, which ranges from medical-grade devices like Empatica E4 to various consumer-grade devices, sampling frequency, and even methodology for inducing stress, ranging from subjective self-reports to objective task-based labeling.

Despite the proliferation of research and data generation, the field of digital stress monitoring is currently characterized by extreme heterogeneity. Existing datasets differ considerably across several critical dimensions:

- Hardware Specifications: Devices range from high-fidelity medical-grade sensors like the Empatica E4 to various consumer-grade wearables with varying sensor sensitivities.
- Technical Parameters: There is a significant lack of standardization in sampling frequencies and data formats.
- Methodological Frameworks: Researchers utilize a wide array of stress-induction protocols and "ground truth" labeling methods, spanning from standardized laboratory tasks to diverse psychometric scales.

This fragmentation has created a significant "harmonization gap". Most current predictive systems are developed using isolated, small-scale datasets, resulting in "sensor-specific" models that lack generalizability across different devices, populations, or real-world settings. Without a structured approach to data harmonization, the ability to fuse these disparate data sources into a unified diagnostic ecosystem remains unfulfilled.

The primary goal of this research is to systematically map and analyze the existing landscape of wearable sensor datasets for stress and anxiety to address the critical fragmentation within the field. By identifying the technical and methodological gaps that currently hinder data integration, this study aims to provide a structured strategy for the development of hardware-agnostic digital biomarkers.

To achieve this, the study employs a rigorous methodology based on the PRISMA 2020 statement for systematic reviews to identify relevant datasets and assess based on a multi-dimensional criteria set, including hardware specifications, physiological modalities, sampling frequencies, and cohort characteristics.

The ultimate objective of this work is to conceptualize a roadmap towards data harmonization to facilitate the creation of more robust, scalable, and clinically relevant monitoring tools for anxiety disorders.

Chapter 2| Background

The scientific exploration of stress has evolved from fundamental physiological observations to digital monitoring. The foundations of this field were established in the early 20th century by Walter Cannon [1], whose work revolutionized the understanding of human emotion. Utilizing early fluoroscopic X-ray technology, Cannon observed that emotional agitation in animals, such as fear or rage, resulted in the immediate cessation of gastrointestinal peristalsis. This led him to introduce the seminal concept of the "fight-or-flight" response, emphasizing the critical role of the sympathetic nervous system (SNS) and the sympathetic-adrenal System (SAS) in maintaining homeostasis. Cannon ultimately proved that emotional stress has a measurable biological cost by showing how the body redirects energy away from non-essential functions to prioritize survival.

Hans Selye subsequently formalized the term "stress" through the General Adaptation Syndrome [2]. This model describes a triphasic physiological response consisting of alarm, resistance, and exhaustion. Selye's research elucidated the pathological effects of chronic exposure to systemic stressors, whereas Cannon's work focused on immediate, acute reactions to emergencies. Selye's later work highlighted the impact of prolonged Hypothalamic-Pituitary-Adrenal (HPA) axis activity and its role in "diseases of adaptation." He provided a definitive link between sustained mental pressure and physical wear-and-tear, a concept now closely associated with allostatic load [3].

Stress assessment was historically confined to laboratory environments and relied on biochemical markers like salivary or serum cortisol, alongside subjective self-reporting [4]. These methods are scientifically validated but remain limited by the "white coat" effect, where the clinical setting itself acts as an extraneous stressor. Furthermore, these "snapshot" measurements fail to capture the dynamic, fluctuating nature of stress in a participant's natural environment.

The shift toward psychophysiology in the late 20th century facilitated the non-invasive measurement of autonomic nervous system (ANS) responses using ECG and EDA. The wearable technology revolution has now bridged the gap between the laboratory and daily life. This emerging paradigm, known as Digital Phenotyping, enables the passive and continuous quantification of stress responses in real-world settings. It effectively translates human behavior into a data-driven clinical reality [5,6].

2.1. Stress, anxiety and panic: key concepts and definitions

2.1.1. Stress

According to the World Health Organization, stress is described as a state of mental tension or worry triggered by challenging situations [7]. It is a natural human response that helps individuals cope with threats or difficulties in daily life. Stress can be categorized into two types based on its duration: acute and chronic. Acute stress refers to short-term stress reactions to immediate challenges or threats, activating the body's "fight or flight" response, which includes physiological changes such as elevated heart rate and the release of adrenaline. In contrast, chronic stress results from prolonged exposure to ongoing stressors and can have serious long-term consequences on both physical and mental health, including an increased risk of cardiovascular diseases, anxiety, and depression [8].

Stress may also be classified according to its source or nature, such as psychological, environmental, or physiological stress. It is a common experience that affects everyone to varying degrees [7].

In the Medical Subject Headings (MeSH) system, stress is categorized under *"Stress Disorders, Traumatic, Acute"*, which is defined as a class of stress disorders characterized by dissociative symptoms that appear shortly after experiencing extreme trauma. These symptoms must resolve within one month; otherwise, the condition is classified as post-traumatic stress disorder (PTSD) [9]. The National Institute of Mental Health defines stress as a reaction to an external event, noting that it typically subsides once the triggering situation has passed [10].

2.1.2. Anxiety and Anxiety disorders

Generally, anxiety is a feeling of fear, dread and uneasiness [11]. It is characterized by feeling of tension and physical changes like increased heart rate and sweating. Anxiety can be mild or severe, and it can be a temporary reaction to stress or a more persistent issue. Anxiety can have many forms such as Generalized Anxiety Disorder (GAD), panic disorder and Phobias. In MeSH, Anxiety is defined as '*Feelings or emotions of dread, apprehension, and impending disaster but not disabling as with ANXIETY DISORDERS*' [12]. Based on National Institute of Mental Health, "Anxiety is an internal thing and it is your reaction to stress." [10].

Anxiety disorders refer to a group of mental disorders characterized by feelings of anxiety and fear, including panic disorder, phobias, social anxiety disorder and post-traumatic stress disorder (PTSD) [13]. In simple terms, we could say that anxiety disorders are stress that continues after the stressor is gone. In MeSH, anxiety disorders are defined as *'Persistent and disabling ANXIETY'* [14]. Based on National Institute of Mental Health, Anxiety disorders involve occasionally fear that doesn't go away and could lead to severe problems. Types of anxiety disorders are panic disorder and social anxiety disorder [15].

2.1.3. Panic

On MeSH tree, panic disorder is a subcategory of anxiety disorders. It is described as a condition characterized by sudden, unexpected panic attacks that usually last several minutes but sometimes can last for hours. These episodes are characterized by overwhelming fear or a sense of impending doom, often accompanied by physical symptoms such as shortness of breath, dizziness, heart palpitations, trembling and a fear of losing control [16]. According to the National Institute of Mental Health, panic disorder involves repeated, unprovoked episodes of intense fear, commonly paired with physical symptoms including chest pain, rapid heartbeat, difficulty breathing, dizziness, or gastrointestinal distress [15].

2.2. Assessment of stress, anxiety and panic

2.2.1. Clinical diagnostic frameworks

The clinical examination of stress and anxiety is a complex and multifaceted process. A thorough clinical interview is the first step in the assessment process in clinical practice. Clinicians search for particular symptoms of panic or generalised anxiety disorders, such as excessive worry, autonomic hyperactivity and a persistent fear of future attacks. Structured interview protocols are typically used to assess a patient's history of exposure to stressors. For this purpose, the "gold standard" is the Life Events and Difficulties Schedule (LEDS) [17]. It entails a thorough, time-consuming interview in which a qualified expert evaluates the significance and context of life events to ascertain whether they pose a long-term threat to the person. More

recent automated approaches, such as the Stress and Adversity Inventory (STRAIN), have been developed to overcome the LEDS's time constraints. The STRAIN maintains high clinical utility while more effectively measuring lifetime stressor exposure through the use of a computer-assisted structure [18].

Recognizing an individual's subjective response to a stressor as an essential part of the clinical examination is often more important than the stressful event itself when determining their health outcome. This was shown in longitudinal studies that examined high-stress populations (e.g. family caregivers), where individuals who had a higher level of psychological burden and emotional strain were significantly more likely to experience a decline in their physical health and die sooner than individuals with less psychological burden and/or emotional strain [19]. As such, a clinical diagnostic evaluation is essentially a comprehensive synthesis of the individual's objective health history, subjective experience with the stressor and their physiological (and/or behavioural) reaction(s) to the stressful event, which will ultimately be the 'ground truth' on which all other forms of monitoring technology will be evaluated against.

To ensure consistency and reliability in this complex evaluation, clinicians rely on established international diagnostic frameworks that provide a standardized language for classifying these observations. The two classification systems governing the diagnosis of stress-related conditions are the DSM-5 and the ICD-11.

The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [20] developed by the American Psychiatric Association (APA), serves as a specialized diagnostic classification system for mental disorders. It provides specific, detailed symptom clusters and duration criteria designed to assist clinicians in identifying distinct disorders, such as generalized anxiety disorder (GAD) or panic disorder. The DSM-5 focuses on the descriptive phenomenology of symptoms, facilitates precise clinical communication and provides a rigorous framework for psychiatric research and the development of targeted treatment protocols.

In contrast, the International Classification of Diseases (ICD-11) [21], maintained by the World Health Organization (WHO), is the global standard for health data, clinical documentation, and statistical reporting. It is not limited to mental health but encompasses the entire spectrum of human morbidity. The ICD-11 provides a common international language that allows member states to monitor health trends and report mortality and morbidity statistics. Regarding stress, the ICD-11 has further refined the classification of "disorders specifically associated with stress," emphasizing the requirement that the clinical presentation be directly linked to an identifiable traumatic event or chronic stressor.

2.2.2. Self-report questionnaires and diagnostic criteria

While structured interviews like the LEDS and STRAIN (discussed in the previous section) focus on the objective exposure to life stressors and adversity, psychometric self-report instruments are used to quantify the individual's subjective psychological response. These instruments are essential in digital health research for providing "labels" for the physiological data captured by sensors. Some commonly used self-report questionnaires are summarized below.

Perceived Stress Scale (PSS) is a self-report with several questions that quantifies perceived stress. It assesses how unpredictable and uncontrollable respondents find their lives. There are many variations of PSS based on the number of questions, for example PSS4 has 4 questions [22]; however, the PSS-10 is widely used because it has superior psychometric properties (reliability and validity) across various populations [23]. As an example, some of the PSS questions are:

1. How often have you been upset because of something that happened unexpectedly?

2. How often have you felt that you were unable to control the important things in your life?
3. How often have you felt nervous and 'stressed'?
4. How often have you felt confident about your ability to handle your personal problems?

Responses are typically recorded on a 5-point Likert scale (0 = Never to 4 = Very Often).

State-Trait Anxiety Inventory (STAI) measures two types of anxiety, state anxiety (S-Anxiety) and trait anxiety (T-Anxiety). S-Anxiety is how someone feels this moment of time and T-Anxiety is how someone generally feels. The questionnaire consists of 40 items, 20 of them for state anxiety and 20 for trait anxiety, and scoring uses a 4-point Likert scale. Scores range from 20 to 80, and higher scores indicate greater anxiety [24].

Positive and Negative Affect Schedule (PANAS) questionnaire measures both positive affect (PA) and negative affect (NA). PA assesses enthusiasm and alertness (10 items in questionnaire), while NA assesses distress, anger and fear (10 items in questionnaire). The score is on a 5-point Likert scale, from 10 to 50 [25].

Self-Assessment Manikins (SAM) is a popular, non-verbal pictorial assessment tool, used to measure emotional response like valence and arousal, in psychological research [26]. Participants rate how they feel about something, like a video, image or situation on valence and arousal scales using the pictorial manikins. Because it uses manikins rather than text, it is particularly useful in cross-cultural research or for reducing the cognitive load of the participant during an experiment. It might also have numeric columns for valence and arousal ranging from 1 to 9.

The Pittsburgh Sleep Quality Index (PSQI), is a self-reported questionnaire that assesses sleep quality and disturbances over a month interval, generating a global score from seven sub scores, such as sleep latency, and duration, to differentiate 'good' from 'poor' sleepers [27].

NASA Task Load Index (NASA-TLX) is a subjective questionnaire that assesses perceived workload across six dimensions – mental demand, physical demand, temporal demand, performance, effort and frustration – to provide an overall workload score [28].

2.2.3. Stress induction methods

While clinical diagnostics and psychometric scales provide insight into a person's baseline state or history, experimental research requires the active elicitation of acute stress or panic under controlled conditions. These induction protocols are essential for "ground truth" generation; they allow researchers to synchronize wearable sensor data with a known, timed stimulus.

In laboratory settings, it is typically to induce stress using validated protocols designed to evoke consistent and quantifiable stress reactions. One of the most well-known tests to induce stress is the Trier Social Stress Test (TSST). The Trier Social Stress Test is a standardized laboratory protocol designed to induce acute psychosocial stress by exposing participants to a socially evaluative and uncontrollable situation, typically involving a mock job interview and an arithmetic task performed in front of an unresponsive panel. This method reliably triggers physiological stress responses, including activation of the hypothalamic–pituitary–adrenal (HPA) axis and increases in cortisol levels [29].

Stroop Color and Word Test (SCWT) is a widely used neuropsychological measure of selective attention, processing speed, and inhibitory control. It presents participants with color words printed in conflicting ink colors, requiring them to inhibit the automatic tendency to read the word and instead name the ink color [30].

This interference reliably increases cognitive load and is commonly used to evaluate executive function and cognitive stress.

An additional common way to induce stress in laboratory settings is the International Affective Picture System (IAPS). IAPS is a standardized database of color photographs developed to provide emotionally evocative stimuli for experimental research. The images are normed on dimensions of valence (pleasantness), arousal (physiological activation) and dominance (perceived control) [31].

2.2.4. Stress, anxiety and panic biomarkers

While clinical diagnostic frameworks and psychometric scales provide essential snapshots of an individual's mental state, they often rely on retrospective recall and subjective interpretation. Similarly, while laboratory induction protocols offer a controlled environment for observing acute reactions, they may not fully capture the complexity and chronicity of stress as it occurs in daily life. To bridge the gap between these clinical 'gold standards' and the lived experience of the individual, research has turned toward biomarkers. These objective, measurable physiological and behavioral indicators allow for the continuous monitoring of the stress response in both controlled and free-living (real-world) conditions. By quantifying the body's reaction through these biomarkers, we can move from intermittent, subjective assessments to a more granular, real-time understanding of how the human organism navigates environmental demands. To understand how these indicators are captured and interpreted, it is necessary to distinguish between two primary categories: (1) neuroendocrine biomarkers, which reflect hormonal changes in body's chemistry, and (2) electrophysiological biomarkers, which capture the immediate activity of the autonomic nervous system.

2.2.4.1. Neuroendocrine biomarkers

The physiological stress response involves the activation of neuroendocrine systems that help the body maintain internal balance. When a stressor is noticed, the sympathetic-adrenomedullary (SAM) system is rapidly activated, leading to the release of the catecholamines neurotransmitters: adrenaline and noradrenaline from the adrenal medulla. These hormones prepare the body for the 'fight or flight' response by increasing heart rate, blood pressure and respiration [32]. Afterwards, the hypothalamic-pituitary-adrenal axis (HPA) is activated. The hypothalamus releases the corticotropin-releasing hormone (CRH) which stimulates the pituitary gland to release adrenocorticotropic hormone (ACTH). Then ACTH acts on the adrenal cortex and promotes the release of cortisol, the primary glucocorticoid responsible for regulating energy balance and the metabolic response to stress [33].

After the onset of acute stress, cortisol levels rise significantly within 10–20 minutes enabling the body to respond effectively to immediate challenges. However, chronic stress can lead to dysregulation of the HPA axis, resulting in persistently elevated or blunted cortisol levels which are associated with adverse health outcomes [34]. During panic attacks, these mechanisms are activated more abruptly leading to short-lived but intense increases in catecholamines and cortisol.

In addition to these primary stress mediators, dehydroepiandrosterone (DHEA) acts as a counter-regulatory hormone that balances the catabolic effects of cortisol and supports adaptation to stress. Higher DHEA levels have been associated with greater resilience and reduced vulnerability to stress-related disorders [35].

Stress hormones can be measured using various biological samples depending on the duration and nature of stress exposure. Blood and saliva are typically used for short-term assessments, while urine and hair provide information on long-term exposure with hair cortisol reflecting cumulative hormone secretion over weeks or

months. More recently, advancements in biosensing technology have enabled the detection of cortisol and related hormones in interstitial fluid and sweat, allowing for continuous, non-invasive monitoring through lab-on-chip or lab-on-skin wearable devices [35]. Overall, hormonal indicators such as cortisol, adrenaline, noradrenaline and DHEA serve as reliable objective biomarkers of stress and anxiety. They complement physiological measures, including Heart Rate Variability (HRV) and EDA, providing a comprehensive understanding of the body's stress response [32–36].

2.2.4.2. Biomarkers based on physiological data

During stress and anxiety, the body activates the autonomic nervous system (ANS), specifically the sympathetic branch, triggering the 'fight-or-flight' response. This reaction occurs within seconds, as the adrenal medulla releases catecholamines (epinephrine and norepinephrine). This leads to an increase in heart rate and respiration, the diversion of blood flow away from the skin and toward the skeletal muscles, and enhanced activity of the eccrine sweat glands. If the stressor persists, other neuroendocrine systems, such as the HPA axis, help maintain alertness and energy, but the immediate physiological changes are dominated by ANS activation. In the case of panic attacks, this response is more intense, typically peaking within minutes and characterized by rapid heartbeat, shortness of breath, and heightened electrodermal activity [7, 8].

Physiological signals are widely employed in the study of stress and anxiety, as they provide measurable indicators of these autonomic and behavioral responses. Commonly used signals include electrodermal activity (EDA), electrocardiography (ECG), blood volume pulse (BVP), heart rate variability (HRV), skin temperature, and accelerometer data. According to MeSH vocabulary, galvanic skin response (GSR) is a change in electrical resistance of the skin caused by an emotional or cognitive event [36]. Electrodermal activity (EDA) is a broader term of GSR that refers to changes in electrical activity of the skin, reflecting the functioning of sweat glands in response to sympathetic activity of the autonomic nervous system [37]. Recent studies confirm that higher-stress levels cause elevated electrodermal activity, making EDA a reliable biomarker for autonomic nervous system activation [38].

Stress further impacts the cardiovascular system by reducing heart rate variability (HRV), suppressing parasympathetic activity, and shifting the autonomic balance toward sympathetic dominance. A systematic review [39], encompassing 37 studies on both acute and chronic stress, demonstrated that stress exposure is consistently associated with a reduction in HRV, particularly in vagal-mediated components, which reflect the activity of the parasympathetic nervous system (the "rest and digest" system). Specifically, the study highlighted declines in RMSSD and HF-HRV metrics. The RMSSD (Root Mean Square of Successive Differences) is a time-domain metric that quantifies the variance between consecutive heartbeats. It is considered the primary marker of parasympathetic activity and is highly sensitive to the rapid, beat-to-beat changes regulated by the vagus nerve. Also, HF-HRV (High-Frequency HRV) is frequency-domain measure (typically ranging from 0.15 to 0.40 Hz) that specifically captures the respiratory sinus arrhythmia—the heart rate's synchronization with the breathing cycle. Like RMSSD, it serves as a reliable indicator of the body's ability to inhibit stress responses and maintain emotional regulation.

The electrical activity of the heart, recorded via an electrocardiogram (ECG), provides valuable information regarding cardiac rhythm and the timing of the heart's internal conduction system [40]. Beyond the frequency of beats, the morphology of the ECG signal offers deeper insights into the stress response. During acute arousal, several key intervals undergo measurable changes: the PR interval (the time for an impulse to travel from the atria to the ventricles) and the QRS complex (the duration of ventricular depolarization) typically

shorten as conduction speed increases. Conversely, the QTc interval, which represents the time required for the heart to electrically "reset" or repolarize (corrected for heart rate), may become prolonged under stress. These metrics, alongside the reduction in HRV, provide a high-resolution signature of the body's physiological reaction to cognitive or emotional demands. A comprehensive study for stress detection using ECG [41,42] confirms that acute mental stress causes changes in ECG such as faster heart rate, shortened PR/QRS intervals, prolonged QTc and lower heart rate variability HRV.

Accelerometer signals in all three axes provide behavioral context that enhances the reliability of stress detection. Most modern wearables utilize 3-axis accelerometers to provide a full picture of motion in 3D space. Accelerometer signals provide essential behavioral context; for instance, identifying whether a spike in heart rate is due to physical exertion or psychological stress (fidgeting, pacing, or restless movements) [43].

The body's response to stress involves significant hemodynamic shifts, most notably the regulation of blood flow to the extremities. When the sympathetic nervous system is activated, it triggers peripheral vasoconstriction, a narrowing of the blood vessels that reduces the volume of blood reaching the microvascular tissue bed. These fluctuations in Blood Volume Pulse (BVP)—the rhythmic expansion and contraction of vessels with each heartbeat—serve as a sensitive physiological marker for stress. Specifically, acute stress typically results in a measurable reduction in BVP amplitude and a decrease in its variability [39,44]. Photoplethysmography (PPG) is used to capture these internal changes non-invasively. A PPG sensor operates by emitting light (often green or infrared) into the skin and measures the light that is either reflected or absorbed by the blood vessels below. Because hemoglobin absorbs light, the amount of light returned to the sensor fluctuates in direct proportion to the volume of blood present. Analysis of this optical BVP signal, allows the calculation of the Inter-Beat Intervals (IBIs), which provides the fundamental data needed to derive Heart Rate Variability (HRV) metrics [45].

Finally, the peripheral vasoconstriction mentioned above not only reduces blood volume but also leads to a decrease in skin temperature [45]. While skin temperature is a useful indicator of sympathetic arousal, it reacts more slowly than EDA or HRV. Furthermore, skin temperature is highly susceptible to environmental ambient temperature and physical activity, meaning it is most effective as a biomarker when analyzed in conjunction with other signals to filter out external noise. In digital health monitoring, skin temperature during a high-arousal event paired with an increase in EDA, can provide a robust multimodal signature for stress or panic.

The translation of these raw physiological signals into actionable insights requires advanced signal processing and Artificial Intelligence (AI) [46]. Raw data from wearable sensors are often noisy due to motion artifacts and environmental factors, necessitating filtration techniques and feature extraction to isolate stress-related patterns. Machine learning (ML) and deep learning (DL) models are then employed to classify these features into different levels of stress or to predict the onset of panic attacks. By training on large-scale datasets, AI algorithms can identify complex, non-linear relationships between multimodal signals—such as the simultaneous rise in EDA and drop in HRV—that are difficult to detect through traditional statistical methods. These computational approaches enable the development of personalized "digital signatures," allowing for predictive interventions and more accurate mental health monitoring in real-world settings.

2.3. Datasets to inform stress, anxiety, and panic research

In recent years, the rapid advancement of wearable technologies has resulted in an increasing number of reviews attempting to unify the fragmented landscape of physiological data and artificial intelligence (AI)

applications in the field of mental health. These reviews are particularly valuable for identifying current trends, technical limitations and gaps in data standardization that are the focus of this thesis.

A significant contribution to this domain was made by Juárez Pegueros & Rodríguez-Arce [47], who performed a systematic review focusing on the physiological data used in stress and anxiety research. Their analysis of 47 datasets reaffirmed the fact that EDA and HR remain the most used physiological modalities. However, the review also highlighted a notable scarcity of long-term, longitudinal datasets collected in "in-the-wild" environments. Furthermore, the study revealed the fact that while the shift towards multimodal data is a positive trend for the domain, the accessibility of the data and the absence of a "gold-standard" for ground truth remain an issue for the comparative analysis of predictive models.

Shifting attention to the clinical and computational dimensions of wearable technology, Abd-alrazaq et al. [48] conducted a scoping review examining the role of Wearable Artificial Intelligence (WAI) in the management of anxiety and depression. While the authors highlight the transformative potential of AI in identifying digital biomarkers for mental health conditions, they also raise the "black box" problem of the current machine learning models. A key conclusion of the study is the need to have diverse populations of participants in the study to avoid bias in the algorithms, a factor often not considered in small-scale studies conducted in the laboratory. The authors argue that for WAI systems to achieve meaningful clinical utility, future research must prioritize model transparency and the inclusion of diverse participant populations.

The broader integration of wearable data within the framework of digital phenotyping was further explored by Choi et al. [49]. Through a systematic review, the study investigated the potential of using passive sensing through both smartphones and wearable devices to monitor stress and mild depression symptoms. The authors categorized digital features into three domains: physical activity, sleep patterns and social interactions. Their findings show the importance of wearable-derived physiological signals in providing the biological foundation necessary for constructing a comprehensive stress profile. The authors highlight the fact that while smartphones can provide contextual information, the ability to continuously monitor the wearer through wearable devices is essential to capture the subtle changes in the autonomic nervous system related to changes in mental health.

Finally, the ecological validity of the datasets is discussed by Baka et al. [50] in their scoping review of digital interventions for young people. Although the authors' main interest lies in the promotion of mental health and the prevention of mental disorders, the information they have on user engagement and ethical data collection is useful for the curation of the datasets. The authors posit that the effectiveness of stress prediction models not only lies in the accuracy of the sensors but also in the relevance and authenticity of the stressors used during data collection. This aligns with the current understanding of the importance of using realistic scenarios in the datasets to ensure the efficacy of digital interventions.

In summary, while these reviews successfully classify commonly used physiological signals and emerging AI methodologies, there remains a clear lack of a systematic framework for harmonizing the diverse and heterogeneous datasets currently used in the field. Most of the existing literature treats datasets as isolated entities, limiting their interoperability and broader applicability. Addressing this gap, the present research proposes a tiered roadmap for integrating heterogeneous datasets into a unified and interoperable diagnostic ecosystem.

Data harmonization is the process of integrating disparate datasets into a unified and consistent format. This procedure is essential in digital health research because wearable sensors often produce data with varying

sampling rates, different measurement units, and inconsistent metadata structures [51,52,53]. Researchers perform harmonization through a series of technical steps, beginning with schema mapping to align data variables and syntactic normalization to standardize file formats. They also apply semantic harmonization to ensure that clinical labels, such as "stress" or "baseline," share a common definition across different studies. This systematic alignment allows for the pooling of data from diverse hardware sources, transforming fragmented raw signals into a large-scale, hardware-agnostic resource for machine learning analysis. By harmonizing these digital biomarkers, the field can move past the limitations of individual, isolated datasets toward more robust and generalizable stress detection models.

Current research is often siloed, with individual datasets failing to provide the diversity and scale required for generalizable clinical models. This thesis addresses these challenges by systematically identifying and categorizing existing datasets and defining a roadmap for integrating existing heterogeneous datasets into a unified and interoperable ecosystem.

Chapter 3| Methodology

This thesis follows a systematic literature review approach to evaluate the current landscape of wearable sensor datasets for stress and anxiety. The research followed the PRISMA 2020 statement [54] to ensure a transparent, reproducible, and structured approach to evidence synthesis. The dataset selection was performed in two steps: first related publications were retrieved and screened to identify relevant datasets. Then, widely used data repositories were searched to identify relevant datasets. The final collection of datasets was screened using a set of selection criteria. The retained datasets were analyzed to extract information related to bibliometrics and also information that can determine the potential and degree of harmonization that can be achieved.

3.1. Data Information and search strategy

A structured search was conducted in PubMed to identify relevant studies reporting datasets on stress, anxiety, and panic collected via wearable devices. PubMed is a free, publicly accessible database of biomedical literature maintained by the USA National Library of Medicine. It provides access to citations, abstracts, and some full-text articles from life sciences and medical journals [55]. The strategy used for PubMed search was structured using a combination of MeSH terms and free-text keywords corresponding to three main concepts: (1) stress or anxiety; (2) wearable devices and (3) machine learning and artificial intelligence. The query aimed to identify studies for stress, anxiety and panic related datasets via wearable devices. The search was limited to the last 10 years, so as to capture signals and datasets produced with current state of the art wearables. The query is presented in Table 1. Systematic/scoping reviews identified via this query were further screened and their reference lists were analyzed to identify additional studies.

Table 1. Query for PubMed search to identify publications using or mentioning data sets of stress, anxiety, and panic physiological biomarkers collected via wearables.

Concept	No	Query
Stress, anxiety and panic	1	("Anxiety Disorders"[Mesh] OR "Stress Disorders, Traumatic, Acute"[Mesh] OR "Stress Disorders, Post-Traumatic"[Mesh] OR "anxiety"[tiab] OR "panic*"[tiab] OR "stress*"[tiab]) AND (y_10[Filter])
Wearables	2	("Wearable Electronic Devices"[Mesh] OR "Wearable*"[tiab] OR "smart watch"[tiab] OR "smart watches"[tiab] OR ("smart"[tiab] AND "watch"[tiab]) OR ("galvanic"[tiab] AND "skin"[tiab]) OR "psychogalvanic"[tiab] OR "electrodermal"[tiab]) AND (y_10[Filter])
Machine Learning and artificial intelligence	3	("Machine Learning"[Mesh] OR "Prediction Methods, Machine"[Mesh] OR "Pattern Analysis, Machine"[Mesh] OR "Neural Networks, Computer"[Mesh] OR "machine learning"[tiab] OR "artificial intelligence"[tiab] OR "deep learning"[tiab] OR "neural network"[tiab] OR "neural networks"[tiab]) AND (y_10[Filter])
Final query	4	(1) AND (2) AND (3)

3.2. Eligibility criteria and literature selection process

The exclusion criteria for articles retrieved from PubMed included:

1. non-English;
2. without abstract;
3. not obtainable;
4. irrelevant based on title and abstract: the subject was not related to anxiety, stress, or panic studies via physiological biomarkers; and
5. no mention of a dataset.

To allow seamless processing, articles were retrieved based on their PMID (the unique PubMed identifier) and subsequently the PubData2XL online service [56] was employed to retrieve rich article metadata via the PubMed application programming interface and export these to a csv (comma separated values) format for easier analysis via Microsoft Excel® tool.

Initially, duplicates and corrigenda or errata were removed. Then, non-English publications and records without an available abstract were excluded, as they did not allow for a reliable preliminary assessment. Subsequently, the title and abstract were analyzed to remove articles irrelevant to the scope of this study, i.e. articles irrelevant to stress, anxiety or panic studies using physiological biomarkers. Finally, the remaining articles were fully analyzed to exclude those that do not report or mention in any way one or more datasets of physiological signals related to stress, anxiety or panic.

3.3. Dataset identification and collection

The final set of articles were screened to identify and extract datasets. Datasets were considered both if they were used in the particular study or if they were only mentioned and cited.

The dataset selection process was based on the following exclusion criteria:

1. Dataset not available, in the case the article did not provide explicit instructions or repositories for dataset retrieval.
2. Data labelling not relevant, i.e. the data relates to stress, anxiety and panic biomarkers, but the dataset does not provide correlation to stress, anxiety or panic. This involved filtering out datasets focused on behavioral activities (e.g. smoking or physical exercise) or general affective states (e.g. happiness or sadness), as these classifications fell outside the specific clinical and physiological scope of this research.
3. Dataset includes only signals that are not compatible with current wearable technology, e.g. magnetoencephalography or electroencephalography.
4. Dataset includes only signals collected via smartphone, while no wearable is employed.

The last two exclusion criteria ensured that the research remained grounded in wearable technology. We excluded any dataset that did not feature at least one signal capable of being captured by a body-worn device, such as a smartwatch, chest belt, or headband. While smartphones are often used in mobile sensing, we

intentionally excluded them from our definition of wearables to focus strictly on sensors that maintain direct or continuous contact with the body for biometric tracking.

3.4. Data extraction for harmonization assessment

A dedicated extraction form was developed to capture two primary categories of information: (1) bibliometric and demographic data, and (2) technical signal characteristics and annotations.

The bibliometric and demographic data includes items such as the year of publication, the study population (e.g., students, clinical patients, or healthy adults), the sample size, and the environmental setting (laboratory vs. real-world/naturalistic). These variables allow for an assessment of the diversity and representativeness of existing stress research.

The technical signal characteristics and annotation captures the hardware devices used, the specific types of physiological signals recorded (e.g., EDA, HRV, BVP, skin temperature), and the corresponding sampling frequencies. Furthermore, the extraction process identified the ground truth labeling methods, such as the specific stress-induction protocols or the psychometric scales used for validation.

Finally, each identified dataset was evaluated to determine its harmonization potential. This assessment focused on the technical compatibility of metadata, signal alignment, and label consistency across different studies. This extraction phase provides the necessary framework for the development of a roadmap for data harmonization, categorizing datasets based on their readiness for integration into a unified diagnostic ecosystem.

Chapter 4| Results

4.1. Study selection

The PubMed query presented in Table 1 was executed on 29 September 2025 and retrieved 483 articles, as shown in Table 2.

Table 2. Results of the query for PubMed search to identify publications using or mentioning data sets of stress, anxiety, and panic physiological biomarkers collected via wearables, on 29 September 2025.

No	Query*	Retrieved Records (on 29.09.2025)
1	("Anxiety Disorders"[Mesh] OR "Stress Disorders, Traumatic, Acute"[Mesh] OR "Stress Disorders, Post-Traumatic"[Mesh] OR "anxiety"[tiab] OR "panic*"[tiab] OR "stress*"[tiab]) AND (y_10[Filter])	875,793
2	("Wearable Electronic Devices"[Mesh] OR "Wearable*"[tiab] OR "smart watch"[tiab] OR "smart watches"[tiab] OR ("smart"[tiab] AND "watch"[tiab]) OR ("galvanic"[tiab] AND "skin"[tiab]) OR "psychogalvanic"[tiab] OR "electrodermal"[tiab]) AND (y_10[Filter])	45,200
3	("Machine Learning"[Mesh] OR "Prediction Methods, Machine"[Mesh] OR "Pattern Analysis, Machine"[Mesh] OR "Neural Networks, Computer"[Mesh] OR "machine learning"[tiab] OR "artificial intelligence"[tiab] OR "deep learning"[tiab] OR "neural network"[tiab] OR "neural networks"[tiab]) AND (y_10[Filter])	346,959
4	(1) AND (2) AND (3)	483

Out of the 483 retrieved records, 43 were systematic/scoping reviews. These were screened based on the exclusion criteria presented in Section 3.2 and 16 of them were retained for reference list analysis, which yielded 324 additional articles. The collective pool of 764 primary research articles was screened as described in Section 3.2 and the process is presented in the PRISMA flow diagram in Figure 1.

From those 764 articles, 74 duplicates were removed. Then, screening based on title and abstract excluded 1 article not in English, 1 article with no abstract, and 328 articles on topics not relevant to this study. From the 360 remaining articles, 144 were not obtainable and 23 did not mention a dataset.

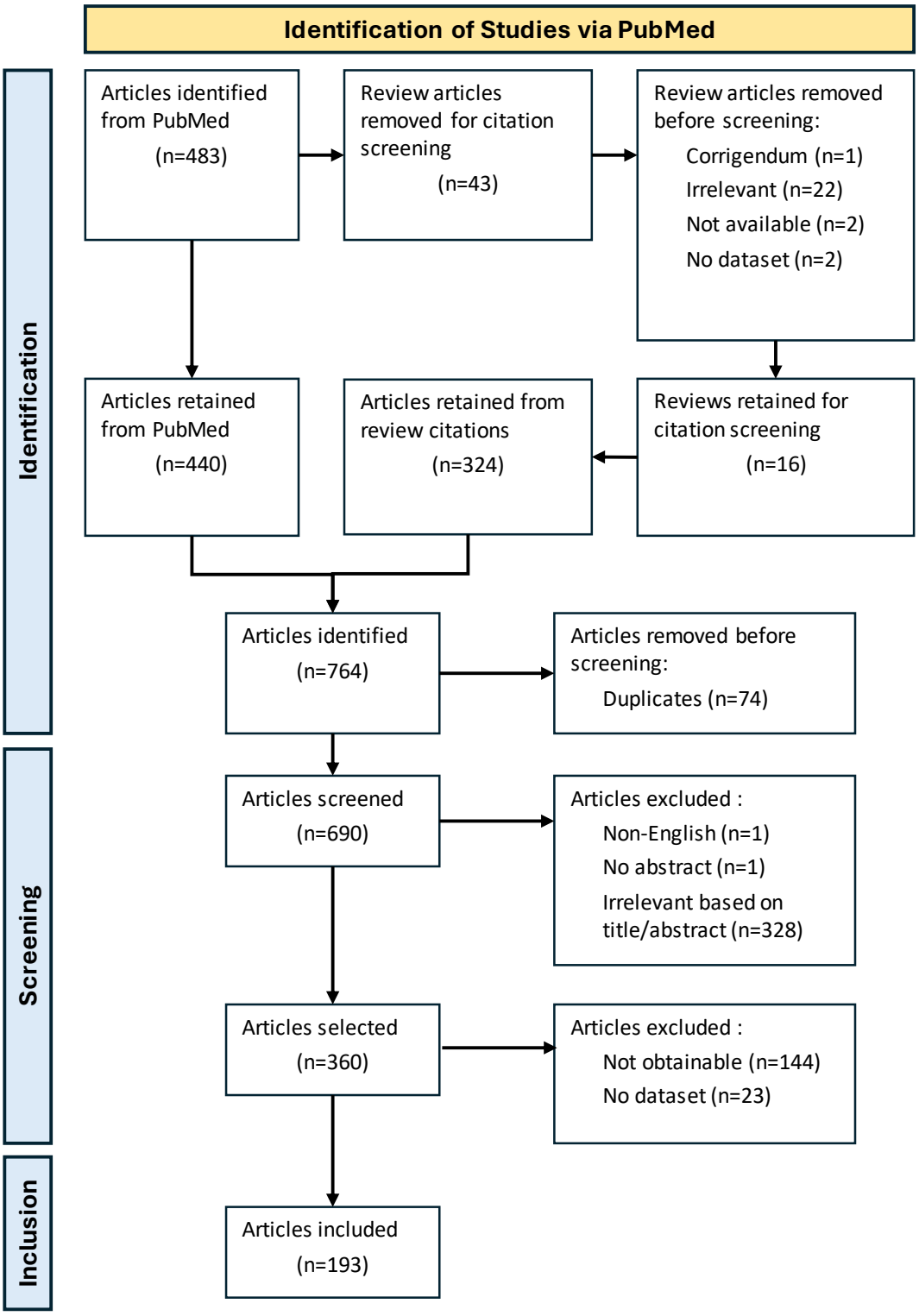

Figure 1. PRISMA flow diagram presenting the article selection process.

The 193 retrieved articles resulted in a total of 201 different data sets for evaluation of eligibility. This difference is due to the discovery of 8 additional data sets in the chosen literature, as the articles referred to more than one independent experimental setting or data source. After removing duplicates, a total of 159 datasets were identified. A refinement process was then initiated to ensure data quality and relevance. Upon attempting data retrieval, 68 datasets were found to be unavailable (e.g., broken links, restricted access, or lack of author response), leaving 91 available for eligibility screening. These remaining datasets were then evaluated against the core exclusion criteria presented in Section 3.3. We excluded datasets with irrelevant labeling (those not mapping to stress, anxiety, or arousal), those consisting exclusively of smartphone data (lacking direct biometric sensing), and those that did not feature at least one potentially wearable-measured signal. This systematic filtering resulted in a final selection of 63 distinct datasets suitable for the objectives of this study. The dataset identification process is shown via a PRISMA 2020 diagram in Figure 2.

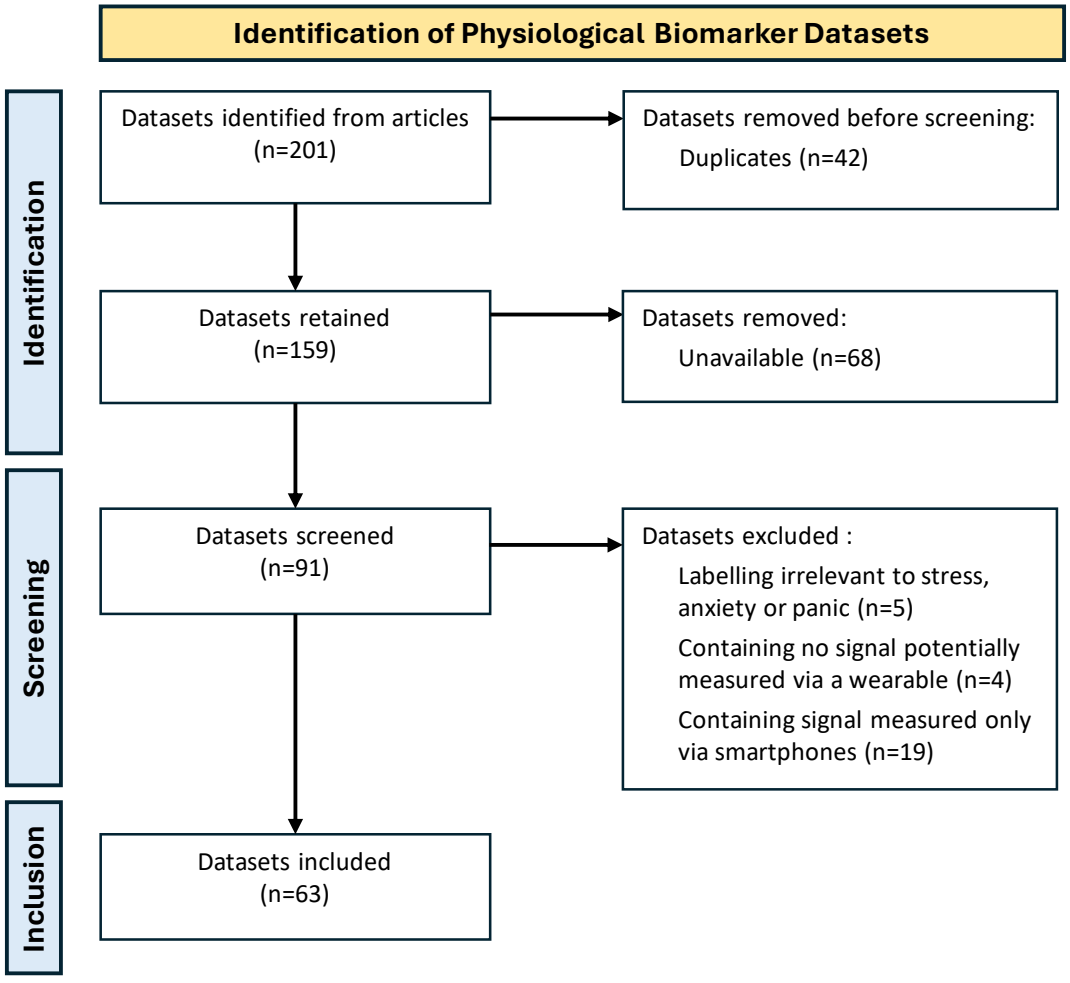

Figure 2. PRISMA flow diagram for the identification of datasets related to physiological biomarkers of stress, anxiety and panic.

The identified datasets and their characteristics are represented in Annex 1. In particular, Table 5 presents datasets based on a unique ID given for this study, together with the dataset name and original citation. Whenever there is no specific name for the dataset, only the citation is included. The table also lists the year

of the dataset collection, and if this is not explicitly mentioned, the year of the publication. Access to the dataset is mentioned as public (when the dataset is available as open access) or as restricted (if the dataset is available only after contacting the authors or on payment). This table also lists information of the type and size of the participants cohort who contributed signals. Information on the signal collection method is included, differentiating between datasets with signals collected in a controlled laboratory environment as opposed to real-world conditions. The method for recording stress levels is included, while the table lists the signals included in each dataset, the wearables used to record theme and the file format in which the data are available.

4.2. Datasets analytics

The 63 selected datasets were analyzed quantitatively to identify the prevailing trends and methodological patterns in current wearable-based stress research. The dataset publication frequency is shown in Figure 3. This analysis shows a transition from sporadic early studies to a modern era of intensive data collection. The substantial increase in datasets from 2018 to 2022 reflects the maturation of wearable sensor technology and the growing demand for multimodal affective datasets. The presence of five datasets for the year 2025 alone demonstrates that real-time stress prediction remains an active research area, continuously evolving with new wearables and diverse cohort descriptions. In terms of dataset accessibility, 52% (n=33) are restricted, requiring formal requests or institutional approval, while the remaining 48% (n=30) are publicly accessible.

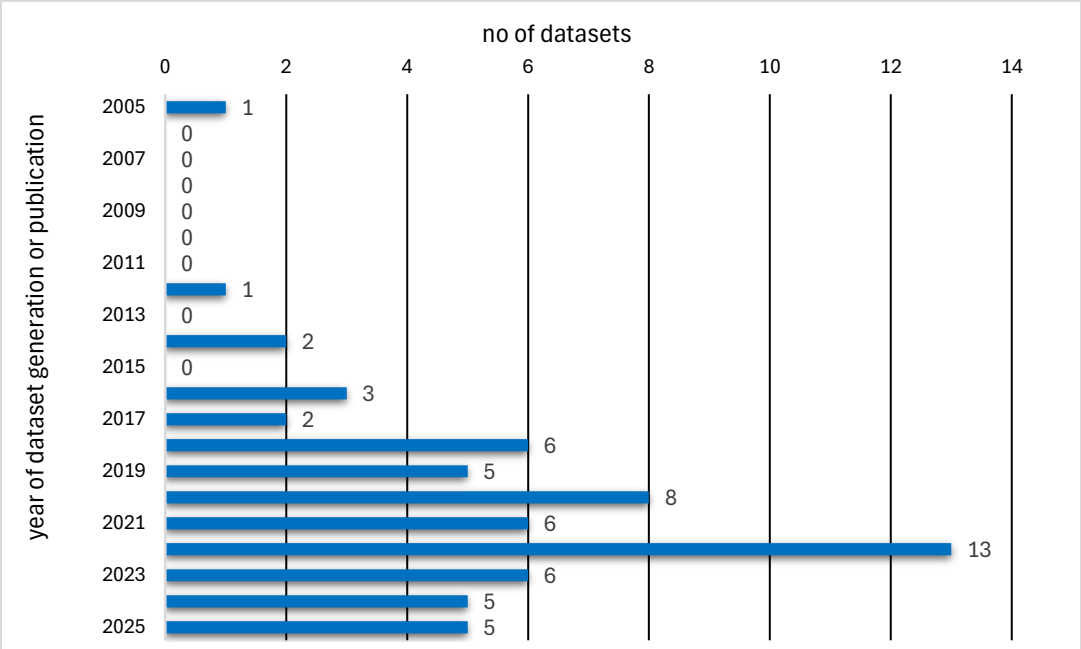

Figure 3. A bar chart of the number of datasets generated or published per year over the last decade.

The analyzed datasets together with the signals included in each one are presented in Table 3. The signals are grouped into six categories based on the system/function they are targeting:

1. skin related:
   - electrodermal activity or galvanic skin response (EDA/GSR);

- integrated skin excursion (ISE), a metric related to EDA and calculated by integrating the area under the skin conductance curve;
- skin temperature (TEMP);

2. heart related:
   - heart rate (HR);
   - heart rate variability (HRV);
   - inter-beat interval (IBI);
   - electrocardiogram (ECG);
   - RR interval in the electrocardiogram (ECG-RR);
3. blood flow related:
   - blood volume pulse (BVP);
   - pulse rate variability (PRV);
   - remote plethysmography (RPPG);
   - peripheral capillary oxygen saturation (SpO_2);
4. respiration related:
   - respiration rate, as indicated by oxygen intake (RESP);
   - breathing rate, as indicated by mechanical chest movement (BR);
   - respiratory inductance plethysmography (RIP);
5. activity related:
   - accelerometer data (ACC);
   - electromyography (EMG);
6. other:
   - electroencephalogram (EEG);
   - eye movement (Eye);
   - sleep quality (Sleep).

The frequency of signal appearance in the analysed datasets (as shown in the last row of Table 3) is plotted in Figure 4.

In the signal classification process, the ECG signal was classified separately from the HR signal to distinguish between raw bio-potential waveforms and derived metrics. While HR is often provided as a pre-processed average through the algorithms of the proprietary devices, the ECG signal is provided as raw data, which is high quality and allows the detailed extraction of complex features like Inter-Beat Intervals and morphology. This distinction is vital for assessing the technical depth of each dataset and its suitability for diverse re-processing and harmonization techniques.

Electrodermal activity (EDA/GSR) is the most prevalent signal, utilized in 63.5% of the studies, followed by heart rate (42.9%) and activity (33.3%). Frequent signals are also skin temperature, electrocardiogram and blood volume pulse, each appearing in 31.7% of the datasets. In contrast, more invasive or complex modalities like EEG and respiration pattern are less frequent.

The number of signals incorporated in the datasets presented in Figure 5. A significant number of datasets (~16%) include only one signal. Most datasets (~59%) include a small number of signals, namely 2-4, while the remaining 25% of datasets include 5 or more signals.

Table 3. Signals included in each one of the analysed datasets. The signals are grouped into 6 categories: skin related, heart related, blood flow related, respiration related and other.

		Skin			Heart					Blood flow				Respiratory			Activity		Other			
ID	Dataset	EDA/GSR	ISE	TEMP	HR	HRV	IBI	ECG	ECG-RR	BVP	PRV	(R)PPG	SpO2	RESP	BR	RIP	ACC	EMG	EEG	Eye	Sleep	No. of Sigals
1	Hongn et al. 2025 - StructuredSession [62]	X		X	X		X			X							X					6
2	SWELL [63]	X						X														2
3	Chen et al. 2023 [64]					X		X		X									X			4
4	Kim H, Kim M et al. 2023 [65]				X			X														2
5	WESAD [66]	X		X			X	X		X				X			X	X				8
6	Ding et al. 2022 [67]	X						X														2
7	Campanella et al. 2023 [68]	X			X	X				X												4
8	Moser et al. 2024 [69]	X		X																		2
9	Elgendi et al. 2022 [70]							X						X								2
10	Xu et al. 2024 - Cares-eskin [71]	X	X	X	X																	4
11	Tutunji et al. 2023 [72]	X		X	X					X							X				X	6
12	Couts, Plans et al. [73]					X				X							X					3
13	Kader et al 2024 [74]	X																X	X			3
14	Healey et al. 2005 -SRAD [75]	X			X			X						X								4
15	Dai et al. 2021 [76]									X							X					2
16	Velmovitsky et al. 2022 [77]					X		X														2
17	Meziati Sabour et al. 2021 - UBFC-Phys [78]	X		X						X	X	X										5
18	Svoren et al. 2019 -Toadstool [79]	X		X	X		X			X							X					6
19	Birjandtalab et al. 2016 - NEURO [80]	X		X	X								X				X					5
20	Haouij et al. 2018 - AffectiveROAD [81]	X		X	X		X			X					X		X					7
21	Hosseini et al. 2022 [82]	X		X	X		X										X					5
22	Mattingly et al. 2019- Tesserae [83]				X																X	2
23	Mevlevioğlu et al. 2024 [84]	X			X							X							X			4
24	Ying Tsai et al. 2025 [85]					X																1
25	Kim HG et al. 2024 [86]	X						X				X							X			4
26	Kenneth Y T Lim et al. 2024 [87]	X																				1
27	PASS [88]	X		X				X		X				X			X		X			7
28	DASPS [89]																		X			1
29	DEAP [90]	X		X						X				X				X	X	X		7
30	Magal et al. 2022 [91]				X												X				X	3
31	EDPMSC [92]																		X			1
32	Sagastibeltza et al. 2023 [93]	X																				1
33	Li et al. 2024 [94]					X		X														2
34	SWEET [95]	X		X	X	X											X					5
35	Ihmig et al. 2020 -Spider [96]	X						X						X								3
36	Iqbal et al. 2022 - Stress-Predict [97]	X			X		X		X	X							X					6
37	Ng et al. 2022 [98]				X	X		X		X												4
38	Başaran et al. 2024 - MoodAware [99]	X				X																2
39	Georgas et al. 2025 [100]	X																				1
40	YAAD [101]	X						X														2
41	EDMSS [102]																		X			1
42	Smerdov et al. 2020 - eSports [103]	X																				1
43	Smets et al. 2018 [104]	X		X	X			X									X					5
44	Said Can et al. 2019 [105]	X			X	X				X							X					5
45	Heussen et al. 2023-IT4Anxiety [106]	X						X														2
46	Quer et al. 2020 -Covid Collab Study [107]				X												X				X	3
47	Shaukat-Jali et al. 2021 [108]	X		X	X																	3
48	Zontone et al. 2020 [109]	X			X	X		X														4
49	Papetti et al. 2025 [110]																X			X		2
50	Seehausen et al. 2025- DiPa [111]	X		X	X	X																4
51	Bin Heyat et al. 2022 [112]					X		X														2
52	Weerasinghe et al. 2023 [113]																		X			1
53	Zhang et al. 2023 [114]	X		X																		2
54	K-EmoCon [115]	X		X	X					X							X		X			6
55	Perpetuini et al. 2021 [116]				X	X				X												3
56	Gjoreski et al. 2017 - Chiron dataset [117]	X			X					X							X					4
57	MMASH [118]					X				X							X				X	4
58	Karl Moser et al. [119]	X		X																		2
59	Hovsepian et al. 2015 - cStress [120]							X								X						2
60	Taamneh et al. 2017 [121]	X			X	X														X		4
61	Said Can et al. 2020 [122]	X		X	X	X											X					5
62	Halim et al. 2020 [123]																		X			1
63	Ghiasi et al. 2020 - BrainBodyStress [124]							X		X									X			3
	No of apperances	40	1	20	27	17	6	20	1	20	1	3	1	6	1	1	21	3	13	3	5	

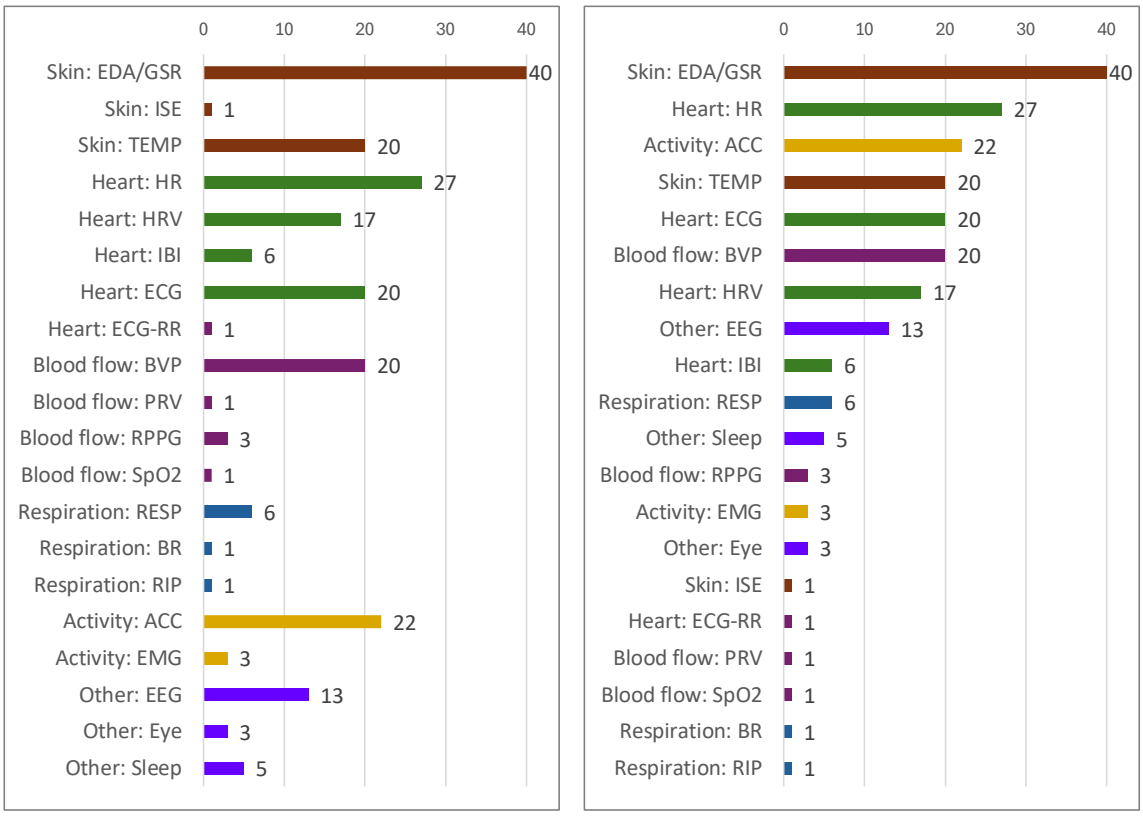

Figure 4. Frequency of signals in the analysed datasets. Left: bar chart with the signals sorted as in Table 3. Right: bar chart with the signals sorted by descending frequency of appearance. Different bar colours indicate different signal groups, as presented in Table 3.

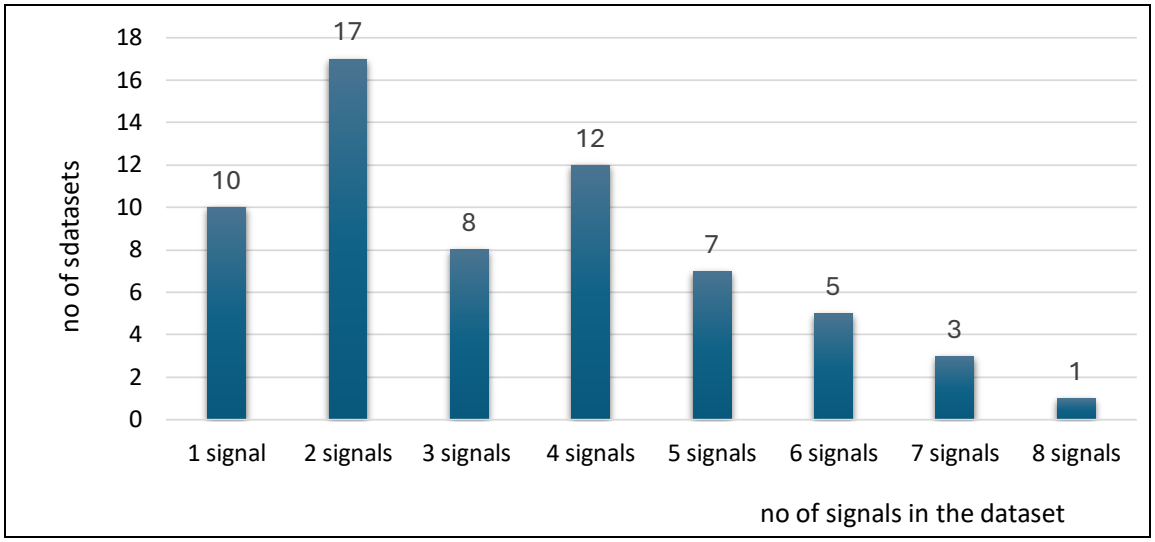

Figure 5. Number of signals incorporated in the datasets.

Table 4 shows a list of commercial wearable devices used in the analyzed datasets. The Empatica E4 [57] dominates the field, appearing in 22 studies, which confirms its status as the standard for research-grade multimodal sensing. The rest of the wearables appear to be used in 4 or less studies each.

Table 4. Commercial wearable devices used to record signals in the analysed datasets.

Wearable	**Provider**	**Frequency**
Empatica E4	Empatica Inc.	22
Zephyr Bioharness	Medtronic Plc	4
Emotiv EPOC / Insight (EEG)	Emotiv Inc	3
Shimmer Sensor	Shimmer Research	3
Samsung Galaxy/Gear Watch	Samsung Electronics	2
NeuroSky MindWave	NeuroSky, Inc.	2
Apple Watch	Apple Inc.	2
Fitbit Watch	Google LLC	2
Garmin Watch	Garmin Ltd.	2
Muse Headband	Interaxon Inc.	2
Polar (H7, Verity Sense)	Polar Electro Oy	2

The stress assessment instruments used for data annotation in the analyzed datasets are presented in Table , while Figure 6 presents a graph showing the frequency of instrument occurrence in the analyzed datasets. Altogether, 27 different questionnaires have been identified, including stress self-reporting scales:

Beck: Beck Anxiety Inventory
BIS/BAS: Behavioral Inhibition System and Behavioral Activation System Scales
CFS: Chalder Fatigue Scale
CSAI: Competitive State Anxiety Inventory
DASS-21: Depression, Anxiety, and Stress Scale (21 Items)
DASS-21-C: DASS-21 Chinese version
DSI: Daily Stress Inventory
EMA: Ecological Momentary Assessment
EPDS: Edinburgh Postnatal Depression Scale
GAD: Generalized Anxiety Disorder Screening Tool
HAM-A: Hamilton Anxiety Rating Scale
LSAS-SR: Liebowitz Social Anxiety Scale - Self-Report
NASA-TLX: NASA Task Load Index
PANAS: Positive and Negative Affect Schedule Scale
PHQ-9: Patient Health Questionnaire-9
PSQI: Pittsburgh Sleep Quality Index
PSS: Perceived Stress Scale, based on the number of items, PSS-10, PSS-14, PSS-5
RAND-36: RAND 36-Item Health Survey

RSME: Rating Scale Mental Effort
SAM: Self-Assessment Manikin Questionnaire
Self-report: generic self-reporting scale
SFS: Social Functioning Scale
SPSQ: Social Phobia Screening Questionnaire
SSSQ: Short Stress State Questionnaire
STAI: State-Trait Anxiety Inventory
SUDS: Subjective Units of Distress Scale
TICS: Trier Inventory for the Assessment of Chronic Stress
VAD: Valence-Arousal-Dominance Model
VAS: Visual Analogue Scale

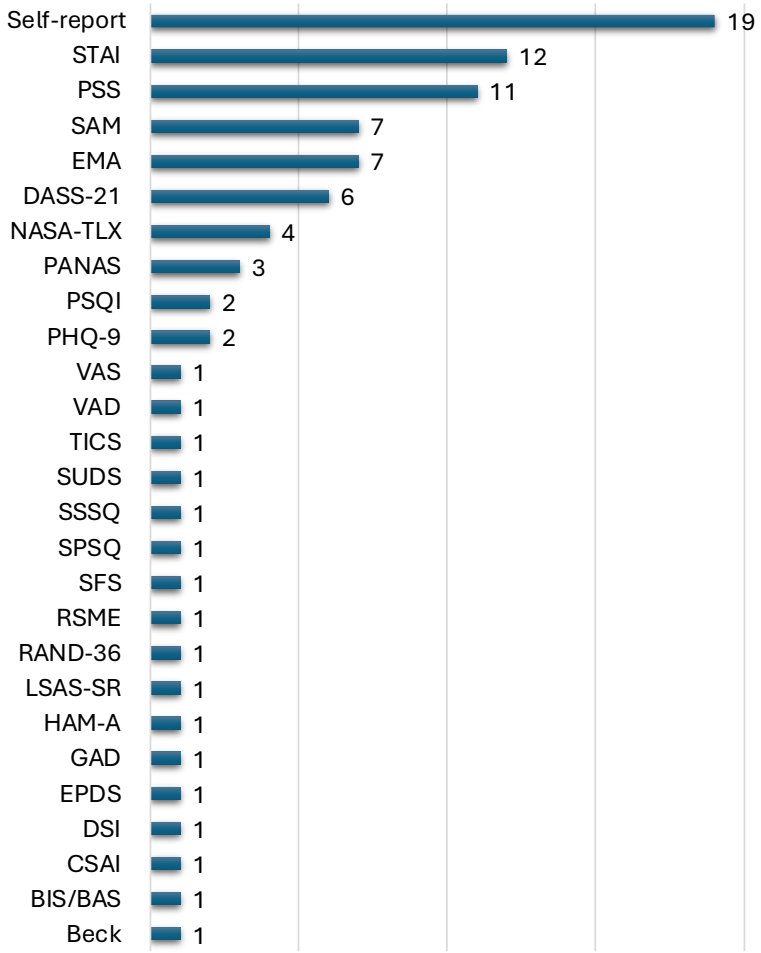

Figure 6. Frequency of stress assessment instruments used in the analysed datasets.

Overall, the number of stress assessment instruments (questionnaires) in each study is shown in Figure 7. Most studies (~51%) use 1 stress assessment instrument, followed by a significant number (~19%) that use 2 stress assessment instruments. There is also a significant number of studies (~16%) that use no stress assessment instrument, rather resort on objective laboratory measurements, e.g. via stress hormone assessment.

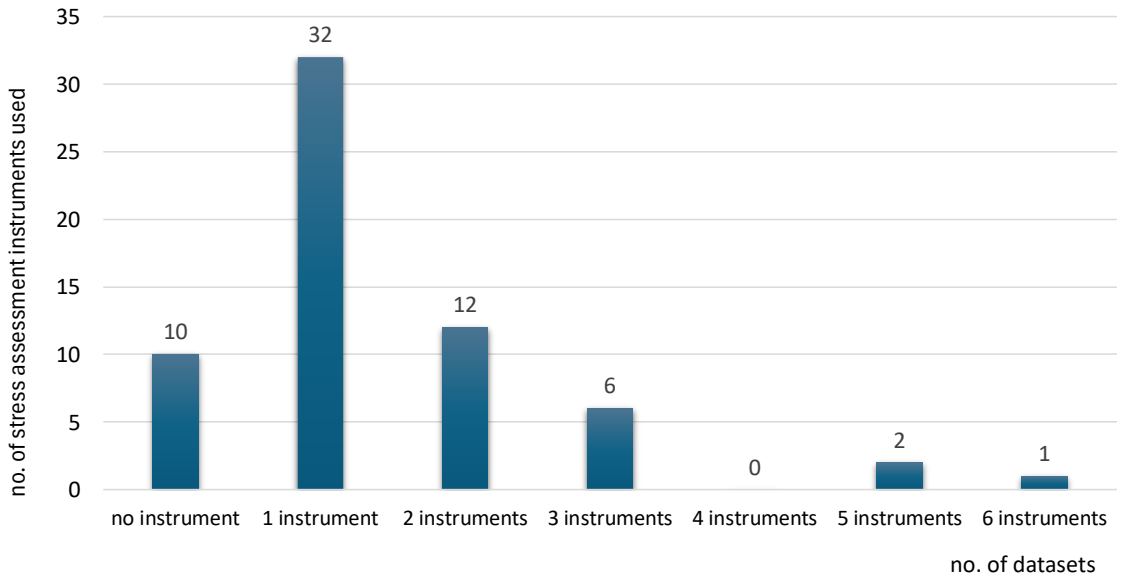

Figure 7. Number of stress assessment instruments (questionnaires) used in the analysed datasets.

Most data sets (69.4%) rely on laboratory environments to acquire data. This concentration is largely due to the high degree of experimental control afforded by laboratory settings, allowing researchers to utilize high-precision equipment and standardized stress induction protocols while minimizing confounding external variables. In contrast, only 27,4% of the analysed datasets were acquired in real-world settings (in-the-wild). These studies are critical for establishing ecological validity, as they capture physiological responses during authentic daily stressors—such as workplace demands or urban driving—despite the inherent challenges of increased signal noise and uncontrolled environmental factors. Finally, a marginal 3.2% of the datasets employ a hybrid methodology. These studies typically initiate data collection using a controlled laboratory baseline for calibration before transitioning participants to long-term monitoring in naturalistic settings. This distribution underscores a significant opportunity for future research to expand upon 'in-the-wild' validation, bridging the gap between controlled proof-of-concept models and robust, real-world mental health applications.

Figure 8 shows the comparative use of some of the most prominently used signals in the analyzed datasets for datasets collected in the laboratory and the real-world settings. In controlled laboratory settings, electrodermal activity (EDA/GSR) is the most often used signal (48% of the datasets. Conversely, in real-world settings a shift toward more robust signals is observed: heart rate (HR) and physical activity (ACC) are used, most often together, to provide indicators of stress normalized to activity. Notably, more invasive or motion-sensitive modalities, such as electroencephalography (EEG) and electrocardiography (ECG), see a drastic reduction in field studies.

Table 6. Stress assessment instruments used for data annotation in the analysed datasets.

ID	Citation	Stress assessment instruments																										No. of Sigals
		Beck	BIS/BAS	CSAI	DASS-21	DSI	EMA	EPDS	GAD	HAM-A	LSAS-SR	NASA-TLX	PANAS	PHQ-9	PSQI	PSS	RAND-36	RSME	SAM	Self-report	SFS	SPSQ	SSSQ	STAI	SUDS	TICS	VAD	VAS	
1	StructuredSession [48]																			X									1
2	SWELL [49]											X						X	X										3
3	Chen et al. 2023 [50]																			X									1
4	Kim H, Kim M et al. 2023 [51]																			X									1
5	WESAD [52]												X						X	X			X	X					5
6	Ding et al. 2022 [53]																							X			X		2
7	Campanella et al. 2023 [54]																												0
8	Moser et al. 2024 [55]																												0
9	Anxiety Dataset 2022 [56]	X																											1
10	Cares-eskin [57]																							X					1
11	Tutunji et al. 2023 [58]						X						X							X									3
12	Couts, Plans et al. [59]				X											X								X					3
13	Kader et al 2024 [60]															X													1
14	SRAD [61]																			X									1
15	Dai et al. 2021 [62]				X											X													2
16	Velmovitsky et al. 2022 [63]				X															X									2
17	UBFC-Phys [64]			X																X									2
18	Toadstool [65]																		X										1
19	NEURO [66]																												0
20	AffectiveROAD [67]																			X									1
21	Multimodal Sensor Dataset [68]																			X									1
22	Tesserae [69]																			X									1
23	AnxietyDataset VR [70]																			X				X					2
24	Ying Tsai et al. 2025 [71]				X																								1
25	VR Stress Interview [72]																											X	1
26	Kenneth Y T Lim et al. 2024 [73]											X							X					X					3
27	PASS [74]											X				X				X									3
28	DASPS [75]									X									X										2
29	DEAP [76]																		X										1
30	Magal et al. 2022 [77]																									X			1
31	EDPMSC [78]															X													1
32	Sagastibeltza et al. 2023 [79]																			X									1
33	Li et al. 2024 [80]													X															1
34	SWEET [81]																			X									1
35	Spider [82]																			X									1
36	Stress-Predict Dataset [83]															X								X					2
37	Ng et al. 2022 [84]						X	X								X													3
38	Mood aware [85]															X													1
39	Georgas et al. 2025 [86]																												0
40	YAAD [87]						X																						1
41	EDMSS [88]																												0
42	Multimodal eSports Dataset [89]																												0
43	Smets et al. 2018 [90]				X		X								X	X	X		X										6
44	Said Can et al. 2019 [91]											X																	1
45	IT4Anxiety [92]																							X	X				2
46	Covid Collab Study [93]								X					X															2
47	Shaukat-Jali et al. 2021 [94]										X											X							2
48	Zontone et al. [95]																												0
49	Papetti et al. [96]																			X									1
50	DiPa [97]						X													X									2
51	Bin Heyat et al. 2022 [98]				X																X								2
52	Weerasinghe et al. 2023 [99]															X													1
53	Zhang et al. 2023 [100]						X																						1
54	K-EmoCon [101]																			X									1
55	Perpetuini et al. 2021 [102]																							X					1
56	Chiron dataset [103]																							X					1
57	MMASH [104]		X			X							X		X									X					5
58	Karl Moser et al. [105]																												0
59	cStress [106]						X																						1
60	Taamneh et al. 2017 [107]																							X					1
61	Said Can et al. 2020 [108]															X													1
62	Halim et al. 2020 [109]																												0
63	BrainBodyStress [110]																												0
	No of apperances	1	1	1	6	1	7	1	1	1	1	4	3	2	2	11	1	1	7	19	1	1	1	12	1	1	1	1	

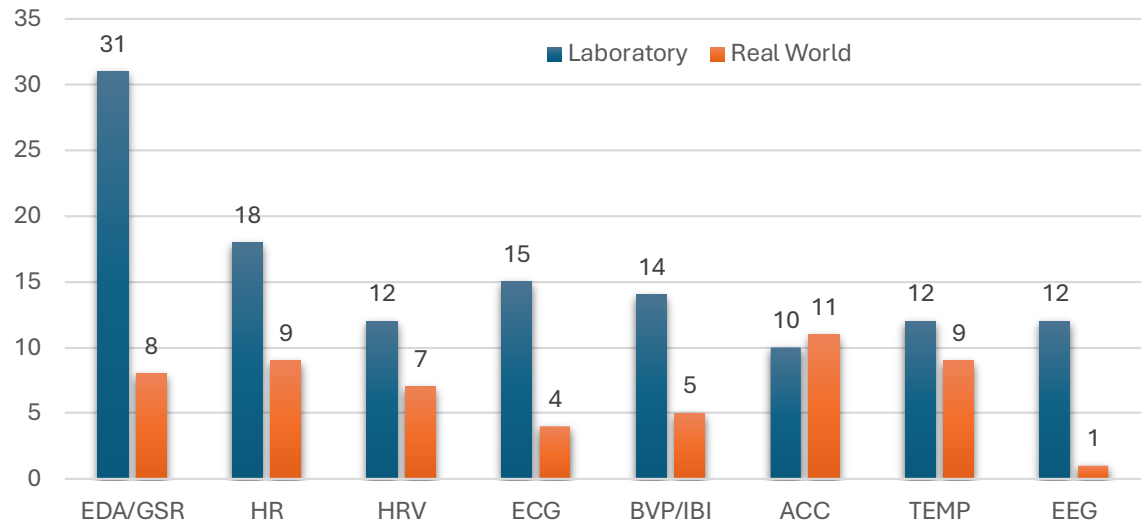

Figure 8. The use of various prominent signals indicating stress in the laboratory (blue) and real-world (orange) settings.

The methods used to trigger stress responses in the analyzed datasets varied and are summarized in Figure 9. Mental and Cognitive tasks are the most common techniques, making up 23.8% of the datasets, with puzzles or logic-based games creating mental demand. Other common lab standards include mental arithmetic and the Stroop color-word test (SCWT). These tests provide reliable stressors that easily sync with physiological data collection.

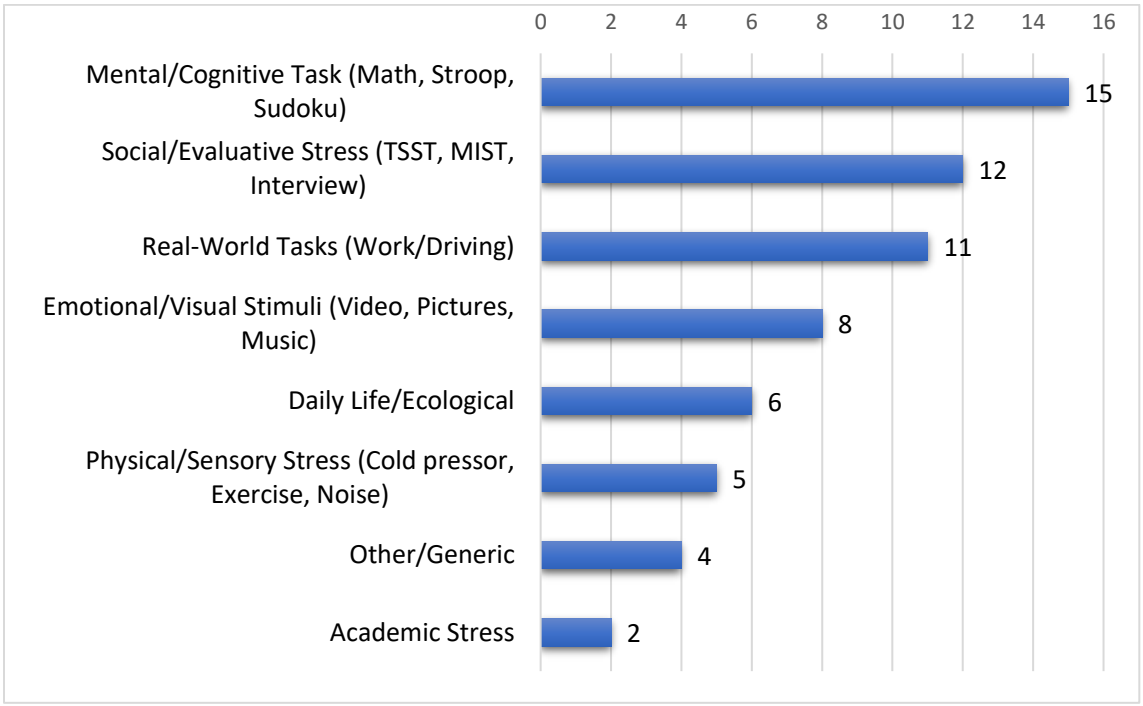

Figure 9. Stress inducing methods used in the analysed datasets.

In addition to cognitive load, social-evaluative stress is also significant. This is mainly achieved through the Trier Social Stress Test (TSST) or mock interviews, used in 19% of the studies to provoke stronger sympathetic nervous system activation. There is also a notable use of affective/media stimuli (12.7%) and real-world scenarios at 17.5%, which include activities like driving or monitoring at work. While lab methods like the TSST offer strong internal validity, the rise in using natural stressors shows a trend towards capturing real-world autonomic responses that more closely reflect everyday human experiences.

The analysed datasets present a diversity regarding participant backgrounds as shown in Figure 10. The largest subgroup consists of students (engaged in 44.4% of the datasets), a common trend in the literature since the academic community serves as the primary and most accessible pool for controlled laboratory experiments. Workers are engaged in 25.4% of the datasets and the general healthy adult population (engaged in 19.0% of the datasets). By including diverse professional groups—such as nurses, commercial drivers, and corporate executives—the reviewed studies capture stress under authentic, high-stakes conditions. This inclusion shifts the analytical focus from "artificial" laboratory-induced triggers to everyday occupational and environmental pressures, providing a more pragmatic view of autonomic reactivity. Finally, a specialized segment of the datasets (11.2%) engages patients or specific clinical groups, such as individuals in palliative care or those diagnosed with clinical anxiety disorders.

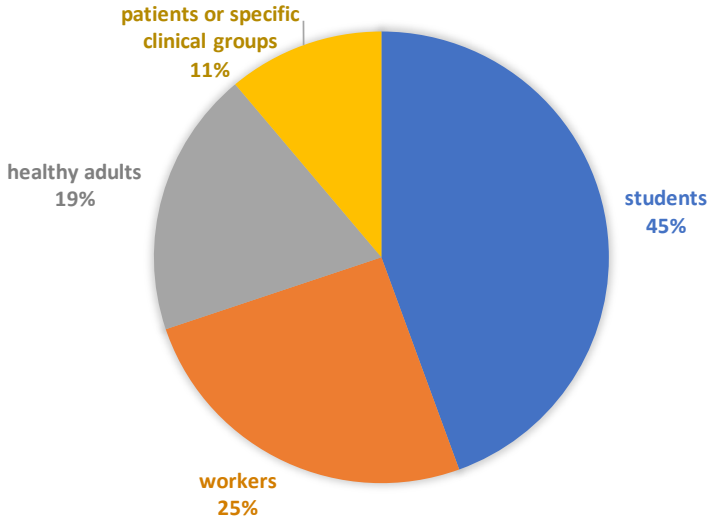

Figure 10. Distribution of the population engaged in the analysed datasets.

A notable limitation identified across the majority of the analyzed datasets was that the participants were mostly healthy adults under the age of 40. This demographic homogeneity, combined with relatively small sample sizes per study, limits the generalizability of the findings to clinical populations, the elderly or individuals with pre-existing conditions, highlighting the need for more diverse and large-scale datasets.

4.3. Potential for data harmonization

Data harmonization is the strategic process of integrating data from different sources – collected using different methodologies, formats and devices – into a unified, consistent and cohesive dataset. According to Cheng et al. [58], harmonization is not a simple merge of files, it is a transformation of data that ensures comparability and analytical utility. In the field of wearable biosensors, harmonization enables us to perform

a cross-analysis of the signals coming from different hardware types (e.g. Empatica E4 vs Shimmer), thus eliminating technical differences. The literature of harmonization procedure [58] suggests a four-stage theoretical framework for successful harmonization:

- Schema Mapping: Identifying common signals/variables (EDA, HR, ACC) across all datasets and aligning them under a standardized naming convention.
- Data Cleaning & Standardization: Correcting data errors and converting units of measurement into a universal system.
- Syntactic Harmonization: Employing signal processing methods such as down sampling or linear interpolation, the varying sample rates can be reconciled, creating a common temporal resolution for all signals.
- Semantic Harmonization: This step aims to harmonize the ground truth labels. This can be achieved, for example, by transforming the heterogeneous stress metrics (such as STAI, Likert 1-5 and binary stress/no stress) into a Unified Stress Index (USI).

Following the four-stage framework outlined above, this research proposes an indicative harmonization plan that adopts a localization of overlap strategy, whereby studies are systematically classified into seven hierarchical tiers according to their technical configuration and physiological compatibility. Rather than treating heterogeneity as a limitation, this approach conceptualizes variability as a structured pathway toward integration. In this way, the framework functions as a practical roadmap for consolidating diverse data sources into a unified and scalable data ecosystem.

At the core of the hierarchy lies Tier 1, which represents the highest level of integrative potential. This tier includes high-fidelity, multimodal datasets collected using the Empatica E4 device and containing the full “tress triad: Electrodermal Activity (EDA), Heart Rate (HR), and Accelerometry (ACC). The strategy to be adopted for the purpose of integrating the data sets is to use the syntactic form of data harmonization, where the sampling rate is aligned to an equivalent resolution, such as 4Hz or 64Hz. The data sets can be integrated to form the primary training set of high consistency for hardware-specific AI models.

Tier 2 extends integration within the same technological ecosystem but focuses specifically on electrodermal monitoring. These datasets, also derived from the Empatica E4 platform, prioritize EDA without necessarily including the complete triad of signals. Harmonization at this level centers on the refinement and validation of skin conductance–based biomarkers. By concentrating on the physiological specificity of EDA, this tier supports the development of robust stress indicators grounded in sympathetic nervous system activity.

Broadening the scope of the research, Tier 3 deals with studies that offer the "stress triad" but with hardware diversity, including Shimmer, BioPac and consumer hardware. For this level, harmonization involves stringent data standardization, including normalization and scaling methods, to account for manufacturer-dependent sensor sensitivity.

Tier 4 addresses datasets that provide raw cardiac waveforms, such as electrocardiogram (ECG) or blood volume pulse (BVP/IBI) signals. At this level, harmonization occurs at the feature-extraction stage. Universal and validated algorithms must be applied to derive Heart Rate (HR) and Heart Rate Variability (HRV) metrics from raw signals, thereby ensuring semantic compatibility with the preprocessed cardiac features available in higher tiers. This process enables meaningful integration across datasets that differ in their degree of preprocessing.

The remaining tiers of the hierarchy deal with special validation issues. Tier 5 deals with univariate electrodermal integration, using datasets that offer only EDA signals to validate the generalizability of the features across different kinds of experimental stressors. Tier 6 deals with neurophysiological validation using an EEG focus, which offers a view into the central nervous system responses that offer a high-level correlation source for the peripheral responses offered by the other tiers. Tier 7 deals with basic cardiac monitoring using datasets that offer only Heart Rate data as external "sparse data" test sets.

The pyramid structure is summarized in Figure 11. The top tier (Tier 1) represents datasets with maximum technical alignment, while the broadening base illustrates the increasing complexity and heterogeneity of data as we move toward sparse or single-signal sources. This hierarchy justifies why different signal processing tactics (e.g., direct merging vs. feature extraction) are applied at different levels.

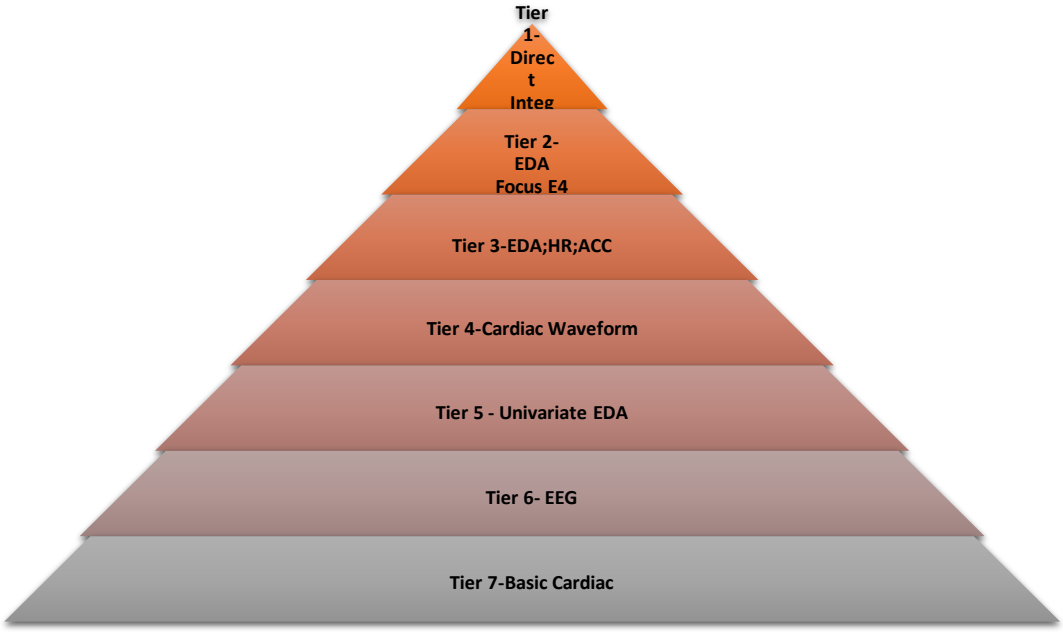

Figure 11. Pyramid of proposed harmonization tiers.

Figure 12 illustrates the quantitative distribution of the 63 datasets across the harmonization hierarchy. Analyzed sets are plotted along the horizontal axis, while the vertical axis shows the various signals included in each dataset presented in the same distinctive color per signal type for all data sets. Similarly, Table shows a resorted list of the datasets, together with the signals they include – only the most commonly encountered signals are considered. From these two alternative visual presentations of the datasets it can be easily derived that signals related to electrodermal activity are the most common, and these datasets provide a first opportunity for direct technical harmonization. Additionally, around 10 datasets include a good number of common signals, thus providing the potential for wider harmonization into a larger dataset with more than one signals.

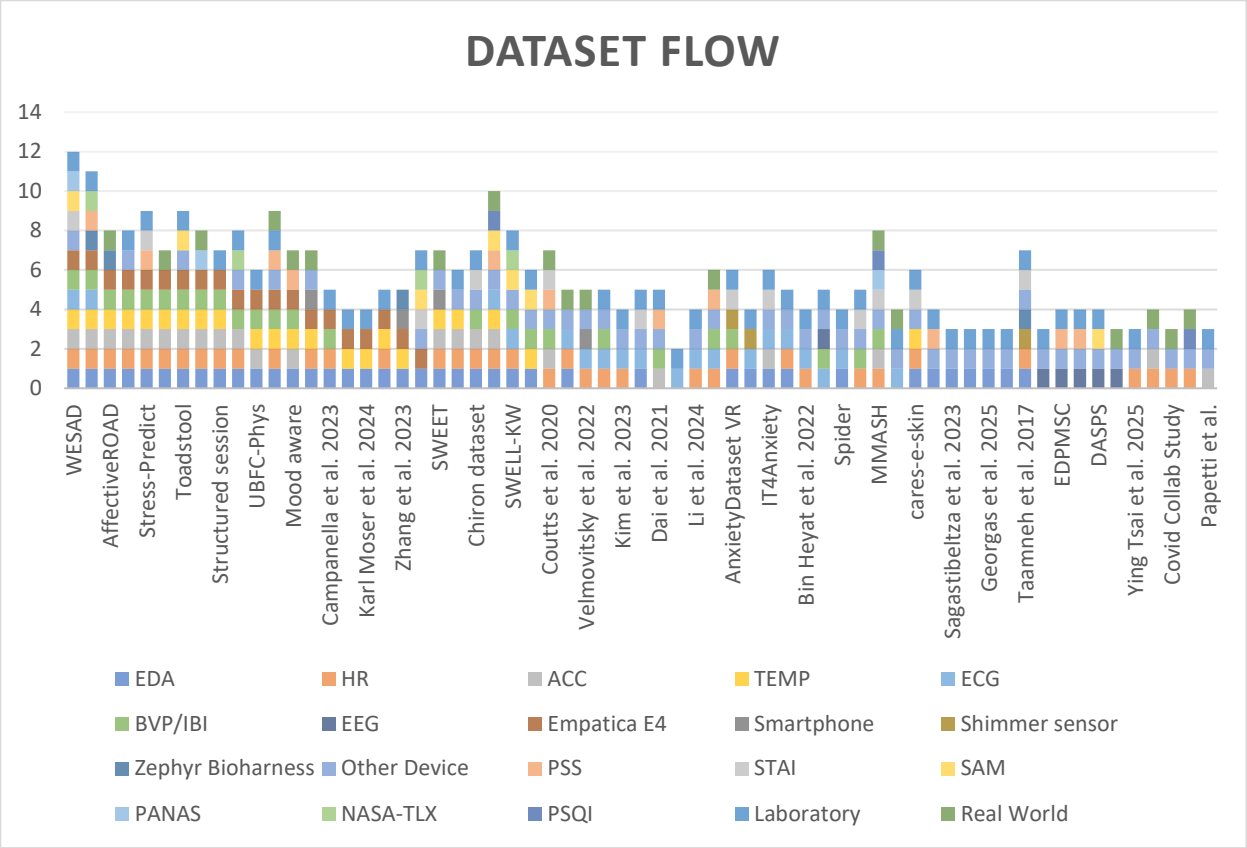

Figure 12. Distribution of the analysed datasets across the harmonization tiers.

Table 7. A list of the datasets resorted to highlight the common signals encountered across the datasets. Only the most commonly encountered signals are considered.

ID	Dataset Name/Source	Empatica E4	EDA/GSR	HR	ACC	BVP/IBI	ECG	EEG
27	PASS [74]	X	X	X	X	X	X	X
21	Multimodal Sensor Dataset [68]	X	X	X	X	X		
36	Stress-Predict Dataset [83]	X	X	X	X	X		
18	Toadstool [65]	X	X	X	X	X		
61	Said Can et al. 2020 [108]	X	X	X	X			
1	StructuredSession [48]	X	X	X	X	X		
11	Tutunji et al. 2023 [58]	X	X	X	X	X		
56	Chiron dataset [103]	X	X	X	X	X		
5	WESAD [52]	X	X	X	X	X	X	
20	AffectiveROAD [67]	X	X	X	X	X		
54	K-EmoCon [101]	X	X	X	X	X		X
26	Kenneth Y T Lim et al. 2024 [73]	X	X					
53	Zhang et al. 2023 [100]	X	X					
17	UBFC-Phys [64]	X	X		X	X		
47	Shaukat-Jali et al. 2021 [94]	X	X	X				
50	DiPa [97]	X	X	X				
8	Moser et al. 2024 [55]	X	X					
7	Campanella et al. 2023 [54]	X	X	X		X		
58	Karl Moser et al. [105]	X	X					
38	Mood aware [85]	X	X		X	X		
44	Said Can et al. 2019 [91]	X	X					
19	NEURO [66]		X	X	X			
34	SWEET [81]		X	X	X			
45	IT4Anxiety [92]		X		X		X	
2	SWELL [49]		X	X		X	X	
51	Bin Heyat et al. 2022 [98]						X	
63	BrainBodyStress [110]	X				X	X	X
35	Spider [82]		X				X	
55	Perpetuini et al. 2021 [102]		X	X		X		
48	Zontone et al. [95]		X	X			X	
43	Smets et al. 2018 [90]		X		X		X	
40	YAAD [87]		X				X	
6	Ding et al. 2022 [53]		X				X	
9	Anxiety Dataset 2022 [56]						X	
14	SRAD [61]		X	X			X	
23	AnxietyDataset VR [70]		X	X		X		X
15	Dai et al. 2021 [62]				X	X		
37	Ng et al. 2022 [84]			X			X	
16	Velmovitsky et al. 2022 [63]						X	
33	Li et al. 2024 [80]						X	
3	Chen et al. 2023 [50]					X	X	X
4	Kim H, Kim M et al. 2023 [51]				X			X
59	cStress [106]						X	
29	DEAP [76]		X			X		X
42	Multimodal eSports Dataset [89]		X					
32	Sagastibeltza et al. 2023 [79]		X					
39	Georgas et al. 2025 [86]		X					
31	EDPMSC [78]							X
52	Weerasinghe et al. 2023 [99]							X
41	EDMSS [88]							X
28	DASPS [75]							X
25	VR Stress Interview [72]		X					X
13	Kader et al 2024 [60]		X					X
62	Halim et al. 2020 [109]							X
22	Tesserae [69]			X				
46	Covid Collab Study [93]			X				
30	Magal et al. 2022 [77]			X	X			
10	Cares-eskin [57]		X	X				
12	Couts, Plans et al. [59]				X	X		
24	Ying Tsai et al. 2025 [71]							
49	Papetti et al. [96]				X			
57	MMASH [104]				X			
60	Taamneh et al. 2017 [107]		X	X				

Chapter 5| Discussion and Conclusions

This systematic review provided a comprehensive mapping of the current landscape in terms of the physiological dataset for the study of stress and anxiety using wearable devices. From the technical and methodological characteristics of 63 unique datasets identified using a rigorous PRISMA-based selection process, several critical conclusions can be drawn regarding the state of the field and the trajectory it is likely to take in the near future.

5.1. Key Findings

The systematic mapping of 63 unique datasets in this study reveals a research landscape for stress and anxiety that is technically advanced yet fundamentally fragmented.

The temporal analysis of dataset publications reveals a field that has transitioned from sporadic exploratory studies into a modern era of intensive data collection. The substantial surge in datasets produced between 2018 and 2022 reflects both the maturation of multimodal wearable technology and a burgeoning academic demand for complex affective computing resources. This momentum persists into the present, with the identification of multiple datasets from 2025 alone, underscoring that real-time stress prediction remains a highly active and evolving research frontier. However, this growth is tempered by significant challenges in data accessibility. More than half of the identified datasets (52%) remain restricted, requiring formal institutional approval or direct requests to authors. This lack of immediate availability presents a hurdle for the rapid validation and cross-comparison of new stress-detection algorithms.

A critical finding of this mapping is the persistent reliance on controlled laboratory environments, which account for nearly 70% of the data analyzed. While these settings offer high experimental control and allow for the use of high-precision equipment and standardized protocols, minimizing external confounding variables, they often lack ecological validity. In contrast, "in-the-wild" studies represent only 27.4% of the landscape. These real-world datasets are essential for capturing authentic physiological responses to everyday stressors, such as workplace demands or urban driving, despite the inherent difficulties of environmental noise. The marginal presence of hybrid methodologies (3.2%), which bridge these two worlds by using laboratory baselines to calibrate long-term naturalistic monitoring, suggests a significant opportunity for future research to transition from proof-of-concept models toward robust, real-world mental health applications.

Technologically, the research landscape is anchored by a core set of physiological signals. Electrodermal Activity (EDA) remains the most prevalent modality, appearing in 63.5% of studies, followed by heart rate and physical activity. While approximately 59% of datasets utilize a moderate range of 2–4 signals, there is a notable trend toward high-dimensionality, with a quarter of the datasets incorporating five or more modalities. Interestingly, the choice of signal appears to be highly dependent on the study environment. In controlled laboratory settings, EDA is the dominant marker; however, in real-world settings, there is a strategic shift toward more robust and motion-resistant signals, specifically the combination of heart rate and accelerometry. This shift allows for the normalization of physiological stress indicators against physical activity, whereas more motion-sensitive or invasive modalities like EEG and ECG see a drastic reduction in field-based applications.

Finally, the methods of stress induction and annotation highlight a tension between standardized reliability and real-world relevance. While cognitive tasks and gamification (27%)—including the Stroop test and mental arithmetic—remain the most common lab standards, there is a growing inclusion of social-evaluative stressors like the Trier Social Stress Test (12.7%) and affective media stimuli (14.2%). For data annotation, the field relies heavily on subjective self-reporting, utilized in 30% of studies, with standardized instruments like the State-Trait Anxiety Inventory (STAI) and the Perceived Stress Scale (PSS) serving as frequent benchmarks. Although approximately 16% of studies opt for objective laboratory measurements such as hormone assessment, the continued rise of real-world scenarios and naturalistic stressors indicates a clear disciplinary trend toward capturing autonomic responses that more closely reflect the everyday human experience.

Finally, the primary finding of this research is the identification of a significant harmonization gap that prevents these disparate data sources from being fused into a unified diagnostic ecosystem. While the rapid proliferation of wearable technology has enabled the massive generation of physiological data, the current environment is characterized by extreme heterogeneity in hardware specifications, sampling frequencies, and stress-induction methodologies. To address this, the central contribution of this thesis is the development of a 7-Tier Harmonization Roadmap, which provides a structured, hierarchical framework for integrating heterogeneous data by accounting for technical interoperability and physiological compatibility across different hardware ecosystems.

5.2. Comparison with current state of the art

These findings align with and significantly expand upon the foundational literature and recent reviews discussed in Chapter 2. Our results confirm that EDA and HR remain the primary modalities, which directly mirrors the systematic review by Juárez Pegueros & Rodríguez-Arce [47], who also noted a scarcity of long-term in-the-wild datasets. The high reliance on student populations observed in this study echoes concerns raised by Abd-alrazaq et al. [48] regarding the "black box" problem and the risk of algorithmic bias in machine learning models.

Furthermore, the importance of wearable-derived signals in providing the biological foundation for stress profiles, as discussed by Choi et al. [49], is supported by the data extracted in this research. While previous reviews by Baka et al. and others successfully classified signals and stressors, they often treated datasets as isolated entities. This thesis moves beyond those observations by operationalizing the theoretical four-stage harmonization framework suggested by Cheng et al. [58] and applying it to a practical roadmap for integration.

While previous reviews by Baka et al. [50] and others successfully classified signals and stressors, they often treated datasets as isolated entities, failing to address how data from one study might complement another. This study fills that gap by operationalizing the theoretical four-stage harmonization framework suggested by Cheng et al. [58] and applying it to a practical roadmap for integration. Furthermore, our findings regarding the lack of standardization in data sharing—where 44.4% of datasets have restricted access—validate the concerns of Schmidt et al. [43] regarding the issue of reproducibility in affective computing.

This work also builds upon the observations of Giannakakis et al. [125] concerning the complexity of emotional state recognition and the need for more granular labeling protocols, by identifying those different studies use vastly different ground-truth labels, ranging from clinical scales like the STAI to simple binary stress markers.

5.3. Strengths and limitations

The primary strength of this work lies in its rigorous and transparent mapping of 63 datasets using a structured PRISMA-based approach. By moving beyond a simple catalog of studies, the 7-Tier Roadmap provides a practical strategy for consolidating diverse data sources—ranging from medical-grade devices to consumer hardware—into a scalable data ecosystem. This approach allows for the inclusion of varied technical configurations, effectively turning the field's inherent heterogeneity into a structured pathway for integration.

However, several methodological limitations must be acknowledged. First, the systematic search was conducted exclusively using PubMed as the primary source of biomedical literature. While PubMed is highly comprehensive for peer-reviewed medical research, this strategy may have excluded datasets that are hosted directly on technical repositories without a corresponding seminal publication indexed in the National Library of Medicine. Specifically, specialized data platforms such as Kaggle, PhysioNet, Figshare and Zenodo were not independently searched. Kaggle is an online platform for data science and machine learning that hosts publicly available datasets and code notebooks, used for research and model development [59]. Figshare is an open-access repository for research outputs, including datasets and figures, allowing researchers to share and cite their work [60]. PhysioNet is a repository of freely accessible physiological and clinical datasets, focusing on biomedical research and signal processing [61]. While some datasets from these repositories were identified indirectly through the literature, a direct search of these platforms might have revealed additional "grey literature" or standalone datasets that are widely used in the machine learning community but less visible in purely clinical databases.

Another significant limitation is the access barrier identified during the collection process. A large portion of the identified datasets (approximately 44.4%) remain under restricted access or require formal institutional approval, which limits their immediate availability for real-time harmonization and cross-analysis. Furthermore, the inherent inconsistency in ground truth labels across the 63 datasets remains a challenge. The variety of subjective measures—ranging from clinical scales like the STAI and PSS to custom Likert-type self-reports—requires a complex process of semantic harmonization that may introduce subjective bias when attempting to map these different metrics to a single Unified Stress Index.

Moreover, the homogeneity of the population in terms of demographics, included of mostly healthy and young participants under the age of 40, limits the applicability of the framework to clinical and elderly populations. Lastly, the absence of metadata on the specific location of the sensors (wrist or fingers) contributes to signal variance, which can be considered as a "black box" in the technical alignment and scalability of the proposed roadmap. Aside from the physical configurations, the lack of standardized metadata, such as inconsistent reporting of sampling rates, data filtering and units of measurements, also represents a significant barrier to data integration.

5.4. Future Work

The findings of this systematic mapping suggest several avenues for future research regarding the methodology and application of data synthesis in the digital biomarker domain. First, while this study established a static snapshot of available datasets, future iterations should aim to develop a dynamic, open-access, living systematic review, resulting also in a digital reference repository. Such a repository would allow researchers to continuously update the harmonization roadmap as new datasets are released, ensuring that the evidence base remains current in a rapidly evolving technological landscape.

Additionally, future work should focus on the technical implementation of the proposed tiers through the development of automated metadata-matching algorithms. These tools could programmatically assess the interoperability of two or more datasets—evaluating sampling rate compatibility, sensor precision, and signal-to-noise ratios—to provide a Harmonization Score before researchers attempt to pool data.

Furthermore, there is a clear need for research into the semantic alignment of ground-truth labels. Future studies could investigate the use of natural language processing (NLP) and ontologies to map disparate psychological assessment tools, such as the STAI, PSS, and various EMA self-reports, onto a standardized, cross-study emotional scale.

Finally, extending this systematic approach to include grey literature and direct searches of technical repositories like PhysioNet and Kaggle would provide a more holistic view of the datasets currently driving the machine learning community, potentially uncovering high-quality resources that are currently underutilized in clinical literature.

5.5. Conclusion

This work provides a foundational strategy for bridging the gap between hardware-restricted experimental research and real-world monitoring systems. By defining a hierarchy for integration, researchers can now pool data from different studies to create larger, more robust training sets for machine learning models. This transition is essential for developing sensor-agnostic digital biomarkers that work reliably across different devices and populations. Ultimately, this facilitates the growth of Digital Phenotyping, turning passive sensor data into a continuous and data-driven clinical reality.

To conclude, while the heterogeneity of the currently available data is a problem, it is also a structured way toward integration. Through systematic harmonization, the currently available landscape of heterogeneous studies can be transformed into a comprehensive and integrated data ecosystem, thereby facilitating the transition toward real-time, non-intrusive mental health monitoring on a global level.

References

[1] Cannon WB. Bodily Changes in Pain, Hunger, Fear and Rage. New York, NY: D. Appleton & Company; 1915.

[2] Selye H. A Syndrome produced by Diverse Nocuous Agents. Nature 138, 32 (1936). https://doi.org/10.1038/138032a0

[3] Selye H. Stress and the general adaptation syndrome. Br Med J. 1950 Jun 17;1(4667):1383-92. https://doi.org/10.1136/bmj.1.4667.1383

[4] Everly GS Jr, Lating JM. A clinical guide to the treatment of the human stress response. 4th ed. New York (NY): Springer; 2019.

[5] Jain SH, Powers BW, Hawkins JB, Brownstein J.S. The digital phenotype. Nature Biotechnology, 33(5), 462-463, 2015. https://doi.org/10.1038/nbt.3223

[6] Torous J, Kiang MV, Lunsford J, Onnela JP. Digital phenotyping: Catching up a mathematical expression to a clinical reality. Schizophrenia Research, 173(3), 124-129, 2016. https://doi.org/10.1016/j.schres.2016.02.043

[7] World Health Organization. Stress. Available from: https://www.who.int/news-room/questions-and-answers/item/stress

[8] Godoy LD, Rossignoli MT, Delfino-Pereira P, Garcia-Cairasco N, de Lima Umeoka EH. A comprehensive overview on stress neurobiology: basic concepts and clinical implications. Front Behav Neurosci, 12:127, 2018. https://doi.org/10.3389/fnbeh.2018.00127

[9] National Library of Medicine (U.S.). MeSH Descriptor Data: Anxiety. 2024. https://www.ncbi.nlm.nih.gov/mesh/68040701

[10] National Institute of Mental Health (NIMH). So Stressed Out: A Fact Sheet. 2023 https://www.nimh.nih.gov/health/publications/so-stressed-out-fact-sheet

[11] MedlinePlus. Anxiety. Available from: https://medlineplus.gov/anxiety.html

[12] National Library of Medicine (U.S.). MeSH Descriptor Data: Stress, Physiological. 2024. https://www.ncbi.nlm.nih.gov/mesh/68001007

[13] Mental disorders [Internet]. Geneva: World Health Organization; 2022 https://www.who.int/news-room/fact-sheets/detail/mental-disorders

[14] National Institute of Mental Health (NIMH). Anxiety Disorders. 2023. https://www.nimh.nih.gov/health/topics/anxiety-disorders

[15] National Library of Medicine (U.S.). MeSH Descriptor Data: Stress, Psychological. 2024. https://www.ncbi.nlm.nih.gov/mesh/68001008

[16] National Library of Medicine (U.S.). MeSH Descriptor Data: Wearable Electronic Devices. 2024. https://www.ncbi.nlm.nih.gov/mesh/68016584

[17] Brown GW, Harris TO. Social Origins of Depression: A Study of Psychiatric Disorder in Women. Free Press; 1978.

[18] Slavich GM, Shields GS. Assessing Lifetime Stress Exposure Using the Stress and Adversity Inventory (STRAIN). Clin Psychol Sci. 80(1):17-27, 2018. https://doi.org/10.1097/PSY.0000000000000534

[19] Schulz R, Beach SR. Caregiving as a risk factor for mortality: the Caregiver Health Effects Study. JAMA, 282(23):2215-9. https://doi.org/10.1001/jama.282.23.22151999

[20] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.

[21] Maercker A, Eberle DJ. Disorders Specifically Associated With Stress in ICD-11. Front Psychol. 4(Spec Issue):e9711, 2022. https://doi.org/10.32872/cpe.97112023

[22] Warttig SL, Forshaw MJ, South J, White AK. New, normative, English-sample data for the short form perceived stress scale (PSS-4). Journal of Health Psychology, 18(12):1617-28, 2013. https://doi.org/10.1177/1359105313508346

[23] Xiao T, Zhu F, Wang D, Liu X, Xi SJ, Yu Y. Psychometric validation of the Perceived Stress Scale (PSS-10) among family caregivers of people with schizophrenia in China. BMJ Open 30;13(11):e076372, 2023. https://doi.org/10.1136/bmjopen-2023-076372

[24] Dalal KS, Chellam S, Toal P. Anaesthesia information booklet: Is it better than a pre-operative visit? Indian Journal of Anaesthesia, 59(8):511, 2015. https://doi.org/10.4103/0019-5049.162998

[25] Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol. 54(6):1063–1070, 1988. https://doi.org/10.1037//0022-3514.54.6.1063

[26] Bynion T-M, Feldner MT. Self-Assessment Manikin, in Zeigler-Hill, Virgil; Shackelford, Todd K. (eds.), Encyclopedia of Personality and Individual Differences, Cham: Springer International Publishing, pp. 1–3, 2017. https://doi.org/10.1007/978-3-319-28099-8_77-1

[27] Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 28(2):193-213, 1989. https://doi.org/10.1016/0165-1781(89)90047-4

[28] Hart SG, Staveland LE. Development of the NASA Task Load Index (TLX): Results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Human Mental Workload. Amsterdam: North Holland; 1988. p. 139-183.

[29] Allen AP, Kennedy PJ, Dockray S, Cryan JF, Dinan TG, Clarke G. The Trier Social Stress Test: Principles and practice. Neurobiol Stress 6:113-126, 2016. https://doi.org/10.1016/j.ynstr.2016.11.001

[30] Scarpina F, Tagini S. The Stroop Color and Word Test. Front Psychol. 8:557, 2017. https://doi.org/10.3389/fpsyg.2017.00557

[31] Lang PJ, Bradley MM, Cuthbert BN. International Affective Picture System (IAPS): Technical Manual and Affective Ratings. NIMH Center for the Study of Emotion and Attention, 1997

[32] Chu B, Marwaha K, Sanvictores T, Awosika AO, Ayers D. Physiology, Stress Reaction. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2023.

[33] Ranabir S, Reetu K. Stress and hormones. Indian J Endocrinol Metab. 15(1):18-22, 2011. https://doi.org/10.4103/2230-8210.77573

[34] Morgan CA 3rd, Southwick SM, Hazlett G, Rasmusson A, Hoyt G, Zimolo Z, Charney D. Relationships among cortisol, DHEA, and subjective responses to stress in humans. Biol Psychiatry 15;56(12):861–868, 2004. https://doi.org/10.1016/j.biopsych.2004.07.014

[35] Kim J, Campbell AS, de Ávila BEF, Wang J. Wearable biosensors for healthcare monitoring. Nat Biotechnol 37(4):389–406, 2019. https://doi.org/10.1038/s41587-019-0045-y

[36] Galvanic Skin Response. In: MeSH – Medical Subject Headings. National Library of Medicine; descriptor data: A change in electrical resistance of the skin, occurring in emotion and in certain other conditions.

[37] Electrodermal activity. In: ScienceDirect Topics. Available from: ScienceDirect Topics: psychology/electrodermal-activity.

[38] Pop-Jordanova N, Pop-Jordanov J. Electrodermal Activity and Stress Assessment. Prilozi (Makedon Akad Nauk Umet Odd Med Nauki), 41(2):5-15, 2020. https://doi.org/10.2478/prilozi-2020-0028

[39] Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. Psychiatry Investig, 15(3):235-245, 2018. https://doi.org/10.30773/pi.2017.08.17

[40] Electrocardiogram (ECG). In: NHS – Tests and treatments. Available from: NHS: electrocardiogram.

[41] Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health, 5:258, 2017. https://doi.org/10.3389/fpubh.2017.00258

[42] Pourmohammadi S, Maleki A. Stress detection using ECG and EMG signals: A comprehensive study. Comput Methods Programs Biomed, 193:105482, 2020. https://doi.org/10.1016/j.cmpb.2020.105482

[43] Schmidt P, Reiss A, Dürichen R, Laerhoven KV. Wearable-Based Affect Recognition-A Review. Sensors (Basel). 19(19):4079, 2019. https://doi.org/10.3390/s19194079

[44] Campanella S, Altaleb A, Belli A, Pierleoni P, Palma L. PPG and EDA dataset collected with Empatica E4 for stress assessment. Data Brief. 53:110102, 2024. https:///doi.org/10.1016/j.dib.2024.110102

[45] Luzzani G, Buraioli I, Guglieri G, Demarchi D. EDA, PPG and Skin Temperature as Predictive Signals for Mental Failure by a Statistical Analysis on Stress and Mental Workload. IEEE Open J Eng Med Biol, 6:248-255, 2024. https://doi.org/10.1109/OJEMB.2024.3515473

[46] Razavi M, Ziyadidegan S, Mahmoudzadeh A, Kazeminasab S, Baharlouei E, Janfaza V, Jahromi R, Sasangohar F. Machine Learning, Deep Learning, and Data Preprocessing Techniques for Detecting, Predicting, and Monitoring Stress and Stress-Related Mental Disorders: Scoping Review. JMIR Ment Health, 11:e53714, 2024. https://doi.org/10.2196/53714

[47] Pegueros JPCJ, Rodríguez-Arce J. Physiological datasets in stress and anxiety research: A systematic review. Comput Biol Med, 108: 107928, 2024. https://doi.org/10.1016/j.bspc.2025.107928

[48] Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, Farooq F, Sheikh J. Wearable Artificial Intelligence for Anxiety and Depression: Scoping Review. J Med Internet Res, 25:e42672, 2023. https://doi.org/10.2196/42672 Erratum in: J Med Internet Res 25:e46233, 2023. https://doi.org/10.2196/46233

[49] Choi A, Ooi A, Lottridge D. Digital Phenotyping for Stress, Anxiety, and Mild Depression: Systematic Literature Review. JMIR Mhealth Uhealth, 12:e40689, 2024. https://doi.org/10.2196/40689

[50] Baka E, Tan YR, Wong BLH, Xing Z, Yap P. A scoping review of digital interventions for the promotion of mental health and prevention of mental health conditions for young people. Oxf Open Digit Health, 3:oqaf005, 2025. https://doi.org/10.1093/oodh/oqaf005

[51] Cheng C, Messerschmidt L, Bravo I, Waldbauer M, Bhavikatti R, Schenk C, Grujic V, Model T, Kubinec R, Barcelo J. A General Primer for Data Harmonization. Sci Data 11, 152, 2024. https://doi.org/10.1038/s41597-024-02956-3

[52] Patel P, Advancements and challenges in data harmonization: A comprehensive review. Journal of Artificial Intelligence and Cloud Computing, 1(2): 1-5, 2022. https://doi.org/10.47363/JAICC/2022(1)234

[53] Iqbal K, Intemann T, Börnhorst C, Zhang J, Aleksandrova K, Approaches for harmonization of biomarker data from multiple studies: a narrative methodological review, AJE Advances: Research in Epidemiology, 1(1): uuaf005, 2025. https://doi.org/10.1093/ajeadv/uuaf005

[54] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 372:n71, 2021. https://doi.org/10.1136/bmj.n71

[55] U.S. National Library of Medicine. PubMed Overview. https://pubmed.ncbi.nlm.nih.gov/about/ Retrieved on Jan 2026.

[56] Rivera Caraballo I. PubData2XML [Internet]. PubData2XL; [cited 2025 Sep 16]. Available from: https://pubdata2xl.com/xml/

[57] E4 wristband: Real-time physiological data streaming and visualization. https://www.empatica.com/research/e4/

[58] Cheng C, Messerschmidt L, Bravo I, Waldbauer M, Bhavikatti R, Schenk C, Grujic V, Model T, Kubinec R, Barceló J. A General Primer for Data Harmonization. Sci Data, 11(1):152, 2024. https://doi.org/10.1038/s41597-024-02956-3

[59] Kaggle. Datasets. https://www.kaggle.com/datasets Retrieved on Jan 2026.

[60] Figshare. About Figshare. https://figshare.com/about Retrieved on Jan 2026.

[61] Goldberger AL, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation, 101(23):e215–e220, 2000. https://physionet.org/about/

[62] Hongn A, Bosch F, Prado LE, Ferrández JM, Bonomini MP. Wearable Physiological Signals under Acute Stress and Exercise Conditions. Sci Data, 28;12(1):520, 2025. https://doi.org/10.1038/s41597-025-04845-9

[63] Koldijk S, Sappelli M, Neerincx MA, Kraaij W. The SWELL knowledge work dataset for stress and working context research. In: Proceedings of the 16th ACM International Conference on Multimodal Interaction (ICMI '14). New York, NY: Association for Computing Machinery; p. 11–18, 2014. https://doi.org/10.1145/2663204.2663257

[64] Chen Q, Lee BG. Deep Learning Models for Stress Analysis in University Students: A Sudoku-Based Study. Sensors (Basel), 23(13):6099, 2023. https://doi.org/10.3390/s23136099

[65] Kim H, Kim M, Park K, Kim J, Yoon D, Kim W, Park CH. Machine learning-based classification analysis of knowledge worker mental stress. Front Public Health. 11:1302794, 2023. https://doi.org/10.3389/fpubh.2023.1302794

[66] Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18); 2018 Oct 16-20; Boulder, CO. New York (NY): Association for Computing Machinery, p. 400-408, 2018. https://doi.org/10.1145/3242969.3242985

[67] Ding Y, Liu J, Zhang X, Yang Z. Dynamic Tracking of State Anxiety via Multi-Modal Data and Machine Learning. Front Psychiatry. 13:757961, 2022. https://doi.org/10.3389/fpsyt.2022.757961

[68] Campanella S, Altaleb A, Belli A, Pierleoni P, Palma L. A Method for Stress Detection Using Empatica E4 Bracelet and Machine-Learning Techniques. Sensors (Basel), 23(7):3565, 2023. https;//doi.org/10.3390/s23073565

[69] Moser MK, Ehrhart M, Resch B. An Explainable Deep Learning Approach for Stress Detection in Wearable Sensor Measurements. Sensors (Basel), 24(16):5085, 2024. https://doi.org/10.3390/s24165085

[70] Elgendi M, Galli V. Ahmadizadeh C. Menon C. Dataset of Psychological Scales and Physiological Signals Collected for Anxiety Assessment Using a Portable Device. Data, 7:132, 2022. https://doi.org/10.3390/data7090132

[71] Xu C, Song Y, Sempionatto JR, Solomon SA, Yu Y, Nyein HYY, Tay RY, Li J, Heng W, Min J, Lao A, Hsiai TK, Sumner JA, Gao W. A physicochemical-sensing electronic skin for stress response monitoring. Nat Electron, 7(2):168-179, 2024. https://doi.org/10.1038/s41928-023-01116-6

[72] Tutunji R, Kogias N, Kapteijns B, Krentz M, Krause F, Vassena E, Hermans EJ. Detecting Prolonged Stress in Real Life Using Wearable Biosensors and Ecological Momentary Assessments: Naturalistic Experimental Study. J Med Internet Res, 25:e39995, 2023. https://doi.org/10.2196/39995

[73] Coutts LV, Plans D, Brown AW, Collomosse J. Deep learning with wearable based heart rate variability for prediction of mental and general health. J Biomed Inform, 112:103610, 2020. https://doi.org/10.1016/j.jbi.2020.103610

[74] Abdul Kader L, Al-Shargie F, Tariq U, Al-Nashash H. One-Channel Wearable Mental Stress State Monitoring System. Sensors (Basel), 24(16):5373, 2024. https://doi.org/10.3390/s24165373

[75] Healey JA, Picard RW. Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions in Intelligent Transportation Systems, 6(2):156-166, 2005. https://doi.org/10.1109/TITS.2005.848368

[76] Dai R, Lu C, Yun L, Lenze E, Avidan M, Kannampallil T. Comparing stress prediction models using smartwatch physiological signals and participant self-reports. Comput Methods Programs Biomed, 208:106207, 2021. https://doi.org/10.1016/j.cmpb.2021.106207

[77] Velmovitsky PE, Alencar P, Leatherdale ST, Cowan D, Morita PP. Using apple watch ECG data for heart rate variability monitoring and stress prediction: A pilot study. Front Digit Health, 4:1058826, 2022. https://doi.org/10.3389/fdgth.2022.1058826

[78] Meziati Sabour R, Benezeth Y, De Oliveira P, Chappé J, Yang F. UBFC-Phys: A Multimodal Database For Psychophysiological Studies Of Social Stress, IEEE Transactions on Affective Computing, 14(1):622, 2021. https://doi.org/10.1109/TAFFC.2021.3056960

[79] Svoren H, Thambawita V, Halvorsen P, Jakobsen P, Garcia-Ceja E, Noori FM, Hammer HL. Toadstool: A Dataset for Training Emotional Intelligent Machines Playing Super Mario Bros. In: Proceedings of the 10th ACM Multimedia Systems Conference (MMSys '19); 2019 Jun 18-21; Amherst, MA. New York (NY): Association for Computing Machinery; p. 301-306, 2019. https://doi.org/10.1145/3339825.3394939

[80] Birjandtalab J, Jafari R, Mazidi M. A non-EEG biosignals dataset for assessment and visualization of neurological status. In: 2016 IEEE International Workshop on Signal Processing Systems (SiPS); 2016 Oct 26-28; Dallas, TX. Piscataway (NJ): IEEE; p. 110-114, 2016. https://doi.org/10.1109/SiPS.2016.27

[81] El Haouij N, Poggi JM, Sevestre-Ghalila S, Ghozi R, Jaïdane M. AffectiveROAD system and database to assess driver's attention. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC '18); 2018 Apr 9-13; Pau, France. New York (NY): Association for Computing Machinery; p. 800-803, 2018. https://doi.org/10.1145/3167132.3167395

[82] Hosseini S, Gottumukkala R, Katragadda S, Bhupatiraju RT, Ashkar Z, Borst CW, Cochran K. A multimodal sensor dataset for continuous stress detection of nurses in a hospital. Sci Data, 9(1):255, 2022. https://doi.org/10.1038/s41597-022-01361-y

[83] Mattingly SM, Gregg JM, Audia P, Bayraktaroglu AE, Campbell AT, Chawla NV, Das Swain V, De Choudhury M, D'Mello SK, Dey AK, Gao G, Jagannath K, Jiang K, Lin S, Liu Q, Mark G, Martinez GJ, Masaba K, Mirjafari S, Moskal E, Mulukutla R, Nies K, Reddy MD, Robles-Granda P, Saha K, Sirigiri A, Striegel A. The Tesserae Project: Large-scale, longitudinal, in situ, multimodal sensing of information workers. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19); 2019 May 4-9; Glasgow, Scotland. New York (NY): Association for Computing Machinery; p. 1-8, 2019. https://doi.org/10.1145/3290607.3299041

[84] Mevlevioğlu D, Tabirca S, Murphy D. Real-Time Classification of Anxiety in Virtual Reality Therapy Using Biosensors and a Convolutional Neural Network. Biosensors (Basel), 14(3):131, 2024. https://doi.org/10.3390/bios14030131

[85] Tsai YY, Chen YJ, Lin YF, Hsiao FC, Hsu CH, Liao LD. Photoplethysmography-based HRV analysis and machine learning for real-time stress quantification in mental health applications. APL Bioeng, 9(2):026103, 2025. https://doi.org/10.1063/5.0256590

[86] Kim HG, Song S, Cho BH, Jang DP. Deep learning-based stress detection for daily life use using single-channel EEG and GSR in a virtual reality interview paradigm. PLoS One, 19(7):e0305864, 2024. https://doi.org/10.1371/journal.pone.0305864

[87] Lim KYT, Nguyen Thien MT, Nguyen Duc MA, Posada-Quintero HF. Application of DIY Electrodermal Activity Wristband in Detecting Stress and Affective Responses of Students. Bioengineering (Basel), 11(3):291, 2024. https://doi.org/10.3390/bioengineering11030291

[88] Parent M, Albuquerque I, Tiwari A, Cassani R, Gagnon JF, Lafond D, Tremblay S, Falk TH. PASS: A Multimodal Database of Physical Activity and Stress for Mobile Passive Body/ Brain-Computer Interface Research. Front Neurosci, 14:542934, 2020. https://doi.org/10.3389/fnins.2020.542934

[89] Baghdadi A, Aribi Y, Fourati R, Halouani N, Siarry P, Alimi AM. DASPS: A Database for Anxious States based on a Psychological Stimulation. arXiv [Preprint]. 2019 Available from: https://arxiv.org/abs/1901.02942

[90] Koelstra S, Muhl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. IEEE Transactions on Affective Computing, 3(1):18-31, 2012. https://doi.org/10.1109/T-AFFC.2011.15

[91] Magal N, Rab SL, Goldstein P, Simon L, Jiryis T, Admon R. Predicting Chronic Stress among Healthy Females Using Daily-Life Physiological and Lifestyle Features from Wearable Sensors. Chronic Stress (Thousand Oaks), 6:24705470221100987, 2022. https://doi.org/10.1177/24705470221100987

[92] Arsalan A, Majid M, Butt AR, Anwar SM. Classification of Perceived Mental Stress Using A Commercially Available EEG Headband. IEEE J Biomed Heal Inform, 23:2257–2264, 2019. https://doi.org/10.1109/JBHI.2019.2926407

[93] Sagastibeltza N, Salazar-Ramirez A, Martinez R, Jodra JL, Muguerza J. Automatic detection of the mental state in responses towards relaxation. Neural Comput Appl, 35(8):5679-5696, 2023. https://doi.org/10.1007/s00521-022-07435-7

[94] Li X, Zou L, Li H. Multilayer Perceptron-Based Wearable Exercise-Related Heart Rate Variability Predicts Anxiety and Depression in College Students. Sensors (Basel), 24(13):4203, 2024. https://doi.org/10.3390/s24134203

[95] Smets E, Rios Velasquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, Cornelis J, Janssens O, Van Hoecke S, Claes S, Van Diest I, Van Hoof C. Dataset of Real-Life Stress Signals From a Large-Scale Study With Wearables (SWEET). Front Digit Health, 3:607533, 2021. https://doi.org/10.3389/fdgth.2021.607533

[96] Ihmig FR, Gogeascoechea H A, Neurohr-Parakenings F, Schäfer SK, Lass-Hennemann J, Michael T. On-line anxiety level detection from biosignals: Machine learning based on a randomized controlled trial with spider-fearful individuals. PLoS One, 15(6):e0231517, 2020. https://doi.org/10.1371/journal.pone.0231517

[97] Iqbal T, Simpkin AJ, Roshan D, Glynn N, Killilea J, Walsh J, Molloy G, Ganly S, Ryman H, Coen E, Elahi A, Wijns W, Shahzad A. Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset. Sensors (Basel), 22(21):8135, 2022. https://doi.org/10.3390/s22218135

[98] Ng A, Wei B, Jain J, Ward EA, Tandon SD, Moskowitz JT, Krogh-Jespersen S, Wakschlag LS, Alshurafa N. Predicting the Next-Day Perceived and Physiological Stress of Pregnant Women by Using Machine Learning and Explainability: Algorithm Development and Validation. JMIR Mhealth Uhealth, 10(8):e33850, 2022. https://doi.org/10.2196/33850

[99] Başaran OT, Can YS, André E, Ersoy C. Relieving the burden of intensive labeling for stress monitoring in the wild by using semi-supervised learning. Front Psychol, 14:1293513, 2024. https://doi.org/10.3389/fpsyg.2023.1293513

[100] Georgas A, Panagiotakopoulou A, Bitsikas G, Vlantoni K, Ferraro A, Hristoforou E. Stress Monitoring in Pandemic Screening: Insights from GSR Sensor and Machine Learning Analysis. Biosensors (Basel), 15(1):14, 2025. https://doi.org/10.3390/bios15010014

[101] Dar MN, Rahim A, Akram MU, Khawaja SG, Rahim A. YAAD: Young Adult's Affective Data Using Wearable ECG and GSR sensors, 2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2), Rawalpindi, Pakistan, pp. 1-7, 2022. https://doi.org/10.1109/ICoDT255437.2022.9787465

[102] Hag A, Handayani D, Pillai T, Mantoro T, Kit MH, Al-Shargie F. Enhancing EEG-Based Mental Stress State Recognition Using an Improved Hybrid Feature Selection Algorithm. Sensors (Basel), 21(24):8370, 2021. https://doi.org/10.3390/s21248370

[103] Smerdov A, Zhou B, Lukowicz P, Somov A. Collection and Validation of Psychophysiological Data from Professional and Amateur Players: a Multimodal eSports Dataset. arXiv [Preprint], 2020. Available from: https://arxiv.org/abs/2011.00958

[104] Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, Cornelis J, Janssens O, Van Hoecke S, Claes S, Van Diest I, Van Hoof C. Large-scale wearable data reveals digital phenotypes for daily-life stress detection. NPJ Digit Med, 1:67, 2018. https://doi.org/10.1038/s41746-018-0074-9

[105] Can YS, Chalabianloo N, Ekiz D, Ersoy C. Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study. Sensors (Basel), 19(8):1849, 2019. https://doi.org/10.3390/s19081849

[106] Heussen S, De Smet J, Badiaga S, Schinköthe T, Lemey L, Op de Beeck T, et al. IT4Anxiety: Empowering patients with anxiety disorders through innovative technologies. Front Psychiatry, 14:1145691, 2023. https://doi.org/10.3389/fpsyt.2023.1145691

[107] Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, Kheterpal V, Topol EJ, Steinhubl SR. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med, 26(10):1661-1664, 2020. https://doi.org/10.1038/s41591-020-1012-3

[108] Shaukat-Jali R, van Zalk N, Boyle DE. Detecting Subclinical Social Anxiety Using Physiological Data From a Wrist-Worn Wearable: Small-Scale Feasibility Study. JMIR Form Res, 5(10):e32656, 2021. https://doi.org/10.2196/32656

[109] Zontone P, Affanni A, Bernardini R, Piras A, Rinaldo R, Formaggia F, Minen D, Minen M, Savorgnan C. Car Driver's Sympathetic Reaction Detection Through Electrodermal Activity and Electrocardiogram Measurements. IEEE Trans Biomed Eng, 67(12):3413-3424, 2020. https://doi.org/10.1109/TBME.2020.2987168

[110] Papetti A, Ciccarelli M, Manni A, Caroppo A, Rescio G. Investigating the Use of Electrooculography Sensors to Detect Stress During Working Activities. Sensors (Basel), 25(10):3015, 2025. https://doi.org/10.3390/s25103015

[111] Seehausen A, Chodan W, Höpfner F, Schneider C, Felser S, Murua Escobar H, Aehnelt M, Junghanss C. Feasibility of Digitally Identifying and Minimizing Stressors in Palliative Care Workplaces by Measuring Stress Continuously for Nurses Through Wearable Sensors (DiPa): Protocol for a Prospective Cross-Sectional Study. JMIR Res Protoc, 14:e63549, 2025. https://doi.org/10.2196/63549

[112] Bin Heyat MB, Akhtar F, Abbas SJ, Al-Sarem M, Alqarafi A, Stalin A, Abbasi R, Muaad AY, Lai D, Wu K. Wearable Flexible Electronics Based Cardiac Electrode for Researcher Mental Stress Detection System Using Machine Learning Models on Single Lead Electrocardiogram Signal. Biosensors (Basel), 12(6):427, 2022. https://doi.org/10.3390/bios12060427

[113] Weerasinghe MMA, Wang G, Whalley J, Crook-Rumsey M. Mental stress recognition on the fly using neuroplasticity spiking neural networks. Sci Rep, 13(1):14962, 2023. https://doi.org/10.1038/s41598-023-34517-w

[114] Zhang Z, Měchurová K, Resch B, Amegbor P, Sabel CE. Assessing the association between overcrowding and human physiological stress response in different urban contexts: a case study in Salzburg, Austria. Int J Health Geogr, 22(1):15, 2023. https://doi.org/10.1186/s12942-023-00334-7

[115] Park S, Aubdieiev I, Ji M, Kim J, Kim J, Choi J, Lee J. K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. Sci Data, 7(1):293, 2020. https://doi.org/10.1038/s41597-020-00630-y

[116] Perpetuini D, Chiarelli AM, Cardone D, Filippini C, Rinella S, Massimino S, Bianco F, Bucciarelli V, Vinciguerra V, Fallica P, Perciavalle V, Gallina S, Conoci S, Merla A. Prediction of state anxiety by machine learning applied to photoplethysmography data. Peer J 9:e10448, 2021. https://doi.org/10.7717/peerj.10448

[117] Gjoreski M, Luštrek M, Gams M, Gjoreski H. Monitoring stress with a wrist device using context. J Biomed Inform, 73:159-170, 2017. https://doi.org/10.1016/j.jbi.2017.08.006

[118] Rossi A, Da Pozzo E, Menicagli D, Tremolanti C, Priami C, Sirbu A, Clifton D, Mariti L, Morelli A, Mauri M, Perciavalle V. Multilevel Monitoring of Activity and Sleep in Healthy People (MMASH): A Complementary Real-World Dataset. Sci Data, 7(1):280, 2020. https://doi.org/10.1038/s41597-020-00617-9

[119] Moser MK, Resch B, Ehrhart M, An Individual-Oriented Algorithm for Stress Detection in Wearable Sensor Measurements, in IEEE Sensors Journal, 23(19):22845-22856, 2023. https://doi.org/10.1109/JSEN.2023.3304422

[120] Hovsepian K, al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment. Proc ACM Int Conf Ubiquitous Comput. 2015:493-504, 2015. https://doi.org/10.1145/2750858.2807526

[121] Taamneh S, Tsiamyrtzis P, Dcosta M, Buddharaju P, Khatri A, Manser M, Ferris T, Wunderlich R, Pavlidis I. A multimodal dataset for various forms of distracted driving. Sci Data, 4:170110, 2017. https://doi.org/10.1038/sdata.2017.110

[122] Can YS, Gokay D, Kılıç DR, Ekiz D, Chalabianloo N, Ersoy C. How Laboratory Experiments Can Be Exploited forMonitoring Stress in the Wild: A Bridge BetweenLaboratory and Daily Life. Sensors (Basel), 20(3):838, 2020. https://doi.org/10.3390/s20030838

[123] Halim ZA, Zaman RF. Stress detection of car drivers using physiological signals and machine learning. J Inf Sci Eng, 36(6):1343-1361, 2020. https://doi.org/10.6688/JISE.202011_36(6).0013

[124] Ghiasi S, Greco A, Barbieri R, Scilingo EP, Valenza G. EEG-Derived Metrics for Mental Stress Assessment Based on Computational Linear and Nonlinear Multimodal Models. Front Neurosci, 14:71, 2020. https://doi.org/10.3389/fnins.2020.00071

[125] Giannakakis G, Grigoriadis D, Giannakaki K, Simantiraki O, Panagiotaki A, Tsiknakis M. Review on Psychological Stress Detection Using Biosignals. IEEE Trans Affect Comput. 2022 Jan-Mar;13(1):440-460. doi: 10.1109/TAFFC.2019.2927337.

ANNEX 1: List of datasets included in the study

Table 7 presents the datasets based on a unique ID given for this study, together with the dataset name and original citation. Whenever there is no specific name for the dataset, only the citation is included. The table also lists the year of the dataset collection, and if this is not explicitly mentioned, the year of the publication. Access to the dataset is mentioned as public (when the dataset is available as open access) or as restricted (if the dataset is available only after contacting the authors or on payment). This table also lists information of the type and size of the participants cohort who contributed signals. Information on the signal collection method is included, differentiating between datasets with signals collected in a controlled laboratory environment as opposed to real-world conditions.

The method for recording stress levels is included, using the following acronyms:

Beck: Beck Anxiety Inventory

BIS/BAS: Behavioral Inhibition System and Behavioral Activation System scales

CFS: Chalder Fatigue Scale

CSAI: Competitive State Anxiety Inventory

DASS-21: Depression, Anxiety, and Stress Scale (21 Items)

DASS-21-C: DASS-21 Chinese version

DSI: Daily Stress Inventory

EMA: Ecological Momentary Assessment

EPDS: Edinburgh Postnatal Depression Scale

GAD: Generalized Anxiety Disorder Screening Tool

HAM-A: Hamilton Anxiety Rating Scale

LSAS-SR: Liebowitz Social Anxiety Scale - Self-Report

MBI: Maslach Burnout Inventory

NASA-TLX: NASA Task Load Index

PANAS: Positive and Negative Affect Schedule Scale

PHQ-9: Patient Health Questionnaire-9

PSQI: Pittsburgh Sleep Quality Index

PSS: Perceived Stress Scale, based on the number of items, PSS-10, PSS-14, PSS-5

RAND-16: RAND 36-Item Health Survey

RSME: Rating Scale Mental Effort

SAM: Self-Assessment Manikin Questionnaire

Self-report: generic self-reporting scale

SFS: Social Functioning Scale

SPSQ: Social Phobia Screening Questionnaire

SSSQ: Short Stress State Questionnaire

STAI: State-Trait Anxiety Inventory

STAI-Y: State-Trait Anxiety Inventory Form Y

SUDS: Subjective Units of Distress Scale

TICS: Trier Inventory for the Assessment of Chronic Stress

VAD: Valence-Arousal-Dominance Model

VAS: Visual Analogue Scale

The table lists the signals included in each dataset; the following acronyms are used:

ACC: Accelerometer data

BR: Breathing Rate

BVP: Blood Volume Pulse

ECG: Electrocardiogram

ECG-RR: RR interval in an ECG

EDA: Electrodermal Activity (often synonymous with GSR)

EEG: Electroencephalogram

EMG: Electromyogram

EOG: Electrooculography, a technique for measuring the resting potential of the retina

GSR: Galvanic Skin Response

HR: Heart Rate

HRV: Heart Rate Variability

IBI: Inter-Beat Interval

ISE: Integrated Skin Excursion (a metric related to EDA and calculated by integrating the area under the skin conductance curve)

MEG: Magnetoencephalography

PRV: Pulse Rate Variability

RESP: Respiration Rate

RIP: Respiratory Inductance Plethysmography, a method to assess breathing patterns

(R)PPG: (Remote) Photoplethysmography

SpO2: Peripheral Capillary Oxygen Saturation

TEMP: Skin Temperature

Table 5. List of analyzed datasets together with information on accessibility, the cohort, method of selection, stress labelling method, signals, and wearables.

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
1	Hongn et al, 2025 - StructuredSession [62]	2025	Public	General population age 18-30 years	36	Laboratory	Self-report (1-10)	EDA BVP IBI HR ACC TEMP	Empatica E4
2	SWELL [63]	2014	Public	Trainees and students	25	Laboratory	NASA-TLX RSME SAM	EDA ECG	Kinet 3D depth cameraMobi device TMSI
3	Chen et al. 2023 [64]	2023	Restricted	Students	30	Laboratory	Custom	BVP HRV ECG EEG	Polar Verity Sense PVS NeuroSky's MindWave Mobile 2 BMD101 EEG
4	Kim H, Kim M et al. 2023 [65]	2023	Restricted	Adults	67	Laboratory	Custom	HR ECG	Samsung Galaxy watch 3
5	WESAD [66]	2018	Public	Students	15	Laboratory	PANAS STAI SAM SSSQ	EDA BVP ECG RESP TEMP	Empatica E4RespiBAN

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
								ACC EMG	
6	Ding et al. 2022 [67]	2022	Restricted	People from college	30	Laboratory	STAI VAD	EDA ECG	GSR Sensor ECG device Electrodes
7	Campanella et al. 2023 [68]	2022	Restricted	Students	29	Laboratory	No	EDA BVP HR HRV	Empatica E4
8	Moser et al. 2024 [69]	2022	Restricted	Adults	28	Laboratory	No	EDA TEMP	Empatica E4
9	Elgendi et al. 2022 - Anxiety Dataset 2022 [70]	2022	Public	Students and staff from Simon Fraser University	19	Laboratory and Real-World	Beck Hamilton	ECG RESP	MP45 BIOPAC Systems
10	Xu et al. 2024 - Cares-eskin [71]	2022	Public	Healthy humans between 23 and 38	10	Laboratory	STAI-Y	GSR ISE HR TEMP Biochemical	CARES electronic skin patch
11	Tutunji et al. 2023 [72]	2019	Restricted	Right-handed, 1st year bachelor students	83	Real-World	PANAS EMA Sleep quality Self reflection	EDA BVP HR ACC	Empatica E4

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
							Custom	TEMP	
12	Couts, Plans et al. [73]	2018	Restricted	Students	Trial1: 68, Trial2: 584	Real-World	PSS STAI DASS-21	HRV BVP ACC	BioBeam Band
13	Kader et al 2024 [74]	2024	Restricted	Right-handed adults	20	Laboratory	PSS	EDA EEG	Electrodes
14	Healey et al. 2005 - SRAD [75]	2005	Public	Drivers	17	Real-world	Self-report	GSR HR ECG EMG RESP	FlexComp system
15	Dai et al. 2021 [76]	2021	Restricted	Washington University in St. Louis	32	Laboratory	PSS DASS-21	BVP ACC	Fossil Gen4 smartwatch
16	Velmovitsky et al. 2022 [77]	2022	Restricted	Locals to the Kitchener-Waterloo region in Ontario	33	Real-World	DASS-21 EMA	HRV ECG	Apple Watch Series 6 Smartphone
17	Meziati Sabour et al. 2021 - UBFC-Phys [78]	2023	Public	Students	56	Laboratory	CSAI	EDA BVP TEMP PRV RPPG	Empatica E4 RGB camera

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
18	Svoren et al. 2019 - Toadstool [79]	2020	Public	General public	10	Laboratory	SAM	EDA BVP IBI HR ACC TEMP	Empatica E4 Camera
19	Birjandtalab et al. 2016 - NEURO [80]	2016	Public	General public	20	Laboratory	--	EDA TEMP ACC HR SpO2	ND
20	Haouij et al. 2018- AffectiveROAD [81]	2018	Public	Drivers	13	Laboratory	Custom	EDA BVP HR IBI BR ACC TEMP	Empatica E4 Zephyr Bioharness 3.0
21	Hosseini et al. 2022 - Multimodal Sensor Dataset for Continuous Stress Detection of Nurses in a Hospital [82]	2020	Public	Nurses	15	Real-World	Custom	EDA HR BVP TEMP ACC	Empatica E4

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
22	Mattingly et al. 2019- Tesserae [83]	2019	Restricted	General public	757	Real-World	PSQI job performance questionnaires intelligence, personality, mood, anxiety, health measures, exercise, sleep, and stress	HR Sleep	Garmin Vivosmart 3
23	Mevlevioğlu et al. 2024 - AnxietyDataset VR [84]	2024	Public	Adults	29	Laboratory	STAI-Y1 Anxiety level test subjective	EDA GSR HR PPG EEG	-Shimmer GSR-MyndPlay MyndBand
24	Ying Tsai et al. 2025 [85]	2025	Restricted	Students	25	Laboratory	DASS-21-C	HRV	Noname Wearable
25	Kim HG et al. 2024 - VR Stress Interview Experiment Data [86]	2022	Public	Young adults 20 years old	30	Laboratory	VAS	GSR ECG EEG PPG	behind-the-ear (BTE) EEG Instrumentation from Biopac System Inc., USA
26	Kenneth Y T Lim et al. 2024 [87]	2024	Restricted	Junior college students	5	Laboratory	STAI SAM NASA-TLX	EDA	Empatica E4 DIY EDA sensor
27	PASS [88]	2020	Public	Adults	48	Laboratory	PSS	EDA	Empatica E4

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
							NASA-TLX Self-reported on physical activity	ECG TEMP BVP RESP EEG ACC	BioHarness 3 Muse Headband
28	DASPS [89]	2019	Not available	Healthy adults	23	Laboratory	SAM HAM-A	EEG	Emotiv EPOC
29	DEAP [90]	2012	Public	Students and staff	32	Laboratory	SAM	EEG GSR BVP EMG TEMP RESP EOG	Biosemi ActiveTwo system
30	Magal et al. 2022 [91]	2022	Restricted	Female participants	140	Real-World	TICS	HR ACC Sleep	Fitbit
31	EDPMSC [92]	2019	Public	General Public	28	Laboratory	PSS	EEG	Muse EEG headband
32	Sagastibeltza et al. 2023 [93]	2023	Restricted	Students	20	Laboratory	Custom	EDA	Biopac EDA -Electrodes
33	Li et al. 2024 [94]	2024	Restricted	University Students	845	Laboratory	PHQ-9	HRV	SmartCardio patch

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
								ECG	
34	SWEET [95]	2018	Restricted	General public	> 1000	Real World	Custom	GSR HR HRV TEMP ACC	Chill Band Wireless EEG Patch
35	Ihmig et al. 2020 - Spider [96]	2017	Public	Adults	80	Laboratory	Custom	GSR RESP ECG	BITalino device
36	Iqbal et al. 2022- Stress-Predict Dataset [97]	2022	Public	Students and staff	35	Laboratory	PSS STAI	EDA BVP HR ECG-RR IBI ACC	Empatica E4
37	Ng et al. 2022 [98]	2022	Restricted	Pregnant women	16	Real-World	PSS EMA EPDS	BVP HR HRV ECG	BioStampRC ECG
38	Başaran et al. 2024- Mood aware [99]	2024	Public	University students	14	Laboratory	PSS-5	EDA HRV	Empatica E4
39	Georgas et al. 2025 [100]	2022	Restricted	People with rapid test for COVID-19	51	Laboratory	--	EDA	GSR Sensor
40	YAAD [101]	2016	Public	Adults	25	Laboratory	EMA	EDA	Shimmer sensor

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
								ECG	
41	EDMSS [102]	2021	Restricted	Students	22	Laboratory	Salivary alpha-amylase levels (AAL)	EEG	EEG headsets
42	Smerdov et al. 2020 - Multimodal eSports Dataset [103]	2021	Public	Players over 22 League of Legends matches	10	Laboratory	--	EDA	No-name wearable
43	Smets et al. 2018 [104]	2018	Restricted	Workers	1002	Real-World	PSS SAM DASS-21 EMA RAND-36 PSQI	EDA HR ECG ACC TEMP Location data Audio signals	Chillband wrist worn Chest patch
44	Said Can et al. 2019 [105]	2018	Restricted	INZVA algorithmic programming contest summer camp participants	21	Laboratory	NASA-TLX	EDA BVP HR HRV ACC	Empatica E4 Samsung Gear
45	Heussen et al. 2023- IT4Anxiety [106]	2025	Restricted	General Public	42	Laboratory	STAI SUDS	EDA ECG	Camera Electrodes Audio signals

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
46	Quer et al. 2020 - Covid Collab Study [107]	2020	Public	General public	>17000	Real-World	GAD PHQ-9	HR ACC Sleep	FitBit Garmin Watch
47	Shaukat-Jali et al. 2021 [108]	2021	Public	Young adults who self identified as shy	13	Laboratory	LSAS-SR SPSQ	EDA HR TEMP	Empatica E4
48	Zontone et al. 2020[109]	2022	Restricted	Drivers	10	Laboratory	--	EDA HR HRV ECG	Wrist wearable ECG Device
49	Papetti et al. 2025 [110]	2025	Restricted	Adults 22 to 39	12	Laboratory	Custom	ACC Eye movement	JINS Meme glasses
50	Seehausen et al. 2025- DiPa [111]	2025	Restricted	Nurses in CPU	12	Real-World	EMA MBI	EDA HR HRV TEMP	Empatica E4 Smartphone NFS Tags Video/Audio
51	Bin Heyat et al. 2022 [112]	2022	Restricted	Researchers	20	Laboratory	DASS-21 SFS CFS	HRV ECG	Smart Tshirt
52	Weerasinghe et al. 2023 [113]	2023	Restricted	Adults	22	Laboratory	PSS	EEG	EEG Quick cap
53	Zhang et al. 2023 [114]	2021	Restricted	Adults	26	Real-World	EMA	EDA TEMP	Empatica E4 Smartphone

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
								Photos	Zehpyr Bioharness
54	K-EmoCon [115]	2020	Public	Adults	32	Laboratory	Custom	EDA BVP HR EEG ACC TEMP Audio signals	Empatica E4 NeuroSky's MindWave Mobile 2
55	Perpetuini et al. 2021 [116]	2021	Restricted	Adults	102	Laboratory	STAI	BVP HR HRV	STMicroelectronics PPG
56	Gjoreski et al. 2017-Chiron dataset [117]	2016	Public	Students	21	Laboratory	STAI	EDA BVP HR ACC	Empatica E4
57	MMASH [118]	2020	Public	General public	22	Real-World	STAI PANAS PSQI BIS/BAS-DSI	BVP HRV ACC Sleep	Polar H7 HR monitor Actigraph Wgt3x-bt
58	Karl Moser et al. [119]	2023	Restricted	Employees and students at the University of Salzburg	16	Laboratory	--	EDA TEMP	Empatica E4

ID	Dataset Name/source	Year	Access	Cohort description	No. of Subjects	Method of signal collection	Method of assessing stress level	Signals collected	Wearables used
59	Hovsepian et al. 2015 - cStress [120]	2014	Restricted	Adults	20	Hybrid	EMA	RIP ECG	Autosense
60	Taamneh et al. 2017 [121]	2017	Public	Drivers	68	Laboratory	STAI Type A/B Personality	EDA HR HRV Eye movement	Thermal Camera Shimmer Sensor FaceLab Zehpyr Bioharness
61	Said Can et al. 2020 [122]	2020	Restricted	Summer school	32	Laboratory	PSS	EDA HR HRV TEMP ACC	Apple Watch Empatica E4
62	Halim et al. 2020 [123]	2019	Restricted	Drivers	50	Laboratory	--	EEG	EMOTIV EPOC+ headset
63	Ghiasi et al. 2020- BrainBodyStress [124]	2020	Public	Adults	18	Laboratory	--	BVP ECG EEG	Empatica E4 EmotivEPOC/Emotiv Insight