Electrodermal Activity (EDA) for Stress Detection: A Comprehensive Survey and Cross-Scenario Benchmark

Gang Liu.
e-mails:gangliu2004@outlook.com

***Abstract*—Continuous stress detection using wearable electrodermal activity (EDA) sensors is a key area in digital health. However, the field is constrained by fragmented public datasets and a lack of standardized evaluation pipelines, which impedes cross-study comparisons and slows progress. Existing surveys either broadly cover multimodal approaches, failing to detail the uniqueness of EDA, or lack systematic rigor. To overcome these barriers, this paper provides the first end-to-end systematic survey and resource consolidation for unimodal EDA-based stress detection. We review the entire technical pipeline from preprocessing to modeling and consolidate 21 public datasets. Crucially, we establish a novel cross-scenario empirical benchmark by evaluating mainstream machine learning and deep learning models on five diverse datasets. Our experiments quantify the core challenges of multi-domain data fusion and cross-domain generalization, revealing the superior performance and transferability of deep learning models, especially when trained on data with high ecological validity. This work provides an essential performance benchmark and empirically-grounded recommendations, aiming to standardize methodologies and accelerate the translation of EDA-based stress detection from laboratory to real-world applications.**

***Index Terms*—Electrodermal Activity, Stress Detection (EDA), Wearable Sensors, Affective Computing, Machine Learning, Deep Learning, Benchmark**

I. INTRODUCTION

With the rapid development of modern society, the definition of health has expanded from the mere absence of physical disease to a comprehensive state of physiological and psychological balance. As a core indicator of an individual's overall well-being, the importance of mental health is increasingly prominent. The fast-paced lifestyle, information overload brought by the internet, and constant social comparison are intensifying the psychological burden on individuals, leading to a frequent occurrence of mental health issues. These problems are not limited to clinically diagnosed mental disorders such as depression or schizophrenia but are more broadly manifested as subclinical symptoms like chronic stress and anxiety, which are silently eroding individuals' quality of life, work efficiency, and interpersonal relationships [1].

Gang Liu has just graduated from Queen Mary University of London, and this work was completed while he was on exchange at the University of Notre Dame.

Stress, as a common psychophysiological response, is essentially an adaptive mechanism produced by the body to cope with internal and external stimuli. According to Hans Selye's classic theory, stress can be divided into two forms: eustress (beneficial stress) and distress (harmful stress) [2]. Eustress can motivate individuals and enhance task performance; for example, moderate exam pressure can promote learning efficiency. However, persistent or excessive distress activates the sympathetic nervous system (SNS), triggering a "fight-or-flight" response and leading to abnormal secretion of stress hormones such as cortisol. This not only interferes with cognitive functions (e.g., distraction and poor decision-making) but may also induce a variety of pathological conditions, including cardiovascular diseases, immune suppression, sleep disorders, and even metabolic disturbances, ultimately causing severe harm to both physical and mental health [3]. According to the latest data from the World Health Organization (WHO), it is predicted that by 2025, more than one billion people worldwide will suffer from mental health problems, among which stress-related issues have become a major factor affecting the working population [4]. In the digital age, remote work models and the amplifying effect of social media further exacerbate these risks, making the development of effective intervention strategies urgently necessary [5].

Therefore, real-time, objective detection of stress has become a critical link in preventing and alleviating mental health problems. Through early identification of high-stress states, individuals can promptly adopt relaxation interventions (such as mindfulness meditation) or seek professional support, thereby effectively preventing the escalation of problems. For example, in daily life scenarios, wearable devices can continuously monitor stress levels, assisting users in optimizing their routines and adjusting their workloads; in high-risk occupations such as driving or healthcare, such monitoring can even trigger alerts to prevent potential accidents [6]. Traditional stress assessment mainly relies on subjective questionnaires (e.g., the Perceived Stress Scale, PSS) or invasive biomarkers (e.g., salivary cortisol detection). Although these methods have a certain degree of reliability, they are generally limited by subjective bias, inconvenient sampling, and poor timeliness, failing to meet the demand for continuous, non-invasive monitoring [7], [8]. In contrast, automated detection based on objective physiological indicators is becoming a mainstream research direction, encompassing various physiological signals such as electrocardiography (ECG) and its derivative, heart rate variability (HRV), electroencephalography (EEG), and photoplethysmography (PPG) [9]. These signals each have their own focus in stress detection: ECG/HRV can highly reflect the balance of the autonomic nervous system (ANS) and is suitable for quantifying the degree of SNS activation [10]; EEG can directly capture cerebral cortex activity, making it particularly suitable for stress research related to cognitive load [11]; and PPG provides non-invasive heart rate indicators by analyzing pulse waves, making it easy to integrate into wearable devices [12]. Although multimodal fusion (e.g., combining ECG, EEG, and PPG) can further improve detection accuracy, it often faces practical challenges such as data synchronization difficulties, high computational complexity, and the need to wear multiple devices [13].

A large body of research has confirmed that relying solely on unimodal EDA for stress detection is also a feasible and efficient solution [14]. EDA is a non-invasive physiological signal that reflects sweat gland activity, which is primarily regulated by the SNS, by measuring dynamic changes in skin conductance [15]. This signal is extremely sensitive to emotional arousal and does not require complex equipment; it can be conveniently collected via wrist-worn sensors (e.g., Empatica E4) and supports real-time analysis. Compared to multimodal methods, the advantages of unimodal EDA are significant: its computational overhead is small, facilitating deployment on resource-constrained edge devices (edge deployment); its technical pathway is simpler, avoiding the heterogeneity of multi-source data and the complexity of fusion algorithms. More importantly, in standard stress-inducing scenarios (e.g., the Trier Social Stress Test, TSST), the correlation between EDA and stress levels is as high as 0.7–0.9, and its performance has been proven to be superior to other single signals or specific combinations [16]–[19]. In summary, with its non-invasive, low-cost, real-time nature, and relative robustness to motion artifacts, EDA has become an ideal choice in the field of daily stress monitoring, providing strong support for early intervention to alleviate distress.

In recent years, EDA-based stress detection has demonstrated broad application value in multiple fields, including safety-critical scenarios such as driving safety, construction sites, and medical monitoring [20]–[22]; as an auxiliary tool to enhance academic performance [23]–[28]; and for work state recognition [29] and human-computer interaction optimization [30]–[32]. These applications not only validate the real-world effectiveness of EDA but also highlight the excellent balance it achieves between computational complexity and detection accuracy [33].

Despite significant progress in existing research, the field still faces several systemic issues that constrain its development. First, **datasets are fragmented and underutilized**. Researchers tend to use a few well-known datasets, leading to a lack of diversity in application scenarios, which severely weakens the cross-scenario generalization ability of the developed models. A systematic review of the vast number of existing datasets is still lacking. Second, **processing pipelines lack uniform standards**. Studies commonly adopt their own signal processing, feature extraction, and validation methods, making it difficult to conduct fair, horizontal comparisons of experimental results, thereby hindering the iteration and advancement of methods within the field. Finally, **the focus of existing surveys is limited**. High-quality surveys mostly focus on multimodal stress recognition, with insufficient depth in the discussion of unimodal EDA; the few surveys that do focus on EDA are lacking in comprehensiveness and systematicity. A detailed comparison between this study and other surveys will be presented in Section II.

To systematically address the aforementioned challenges and promote the standardized development of the field, this paper makes the following core contributions:

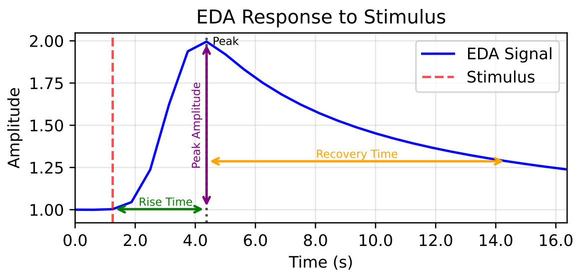
1. **Provide a comprehensive integration of methodologies and a compilation of resources.** This paper presents the first systematic review of the end-to-end pipeline for EDA-based stress detection, covering all stages from data preprocessing and signal decomposition to feature engineering and modeling. We have compiled and analyzed 21 public stress datasets that include EDA, providing a centralized and valuable resource library for future researchers. Furthermore, this paper provides a structured overview of nonlinear methods, as well as unimodal and multimodal approaches in the field, clearly outlining the technical landscape for new researchers.

2. **Establish a much-needed cross-scenario empirical performance benchmark.** To address the problems of non-uniform methods and incomparable results, this paper proposes and implements a standardized evaluation pipeline. Using five datasets covering different scenarios, we conduct a large-scale, cross-scenario horizontal performance evaluation of traditional machine learning models and state-of-the-art deep learning models. This work provides the first empirical benchmark for tackling the two core challenges of multi-scenario data fusion and cross-domain generalization. Based on the experimental results, we offer guiding recommendations for model selection, establishing a performance reference for the future development of the field.

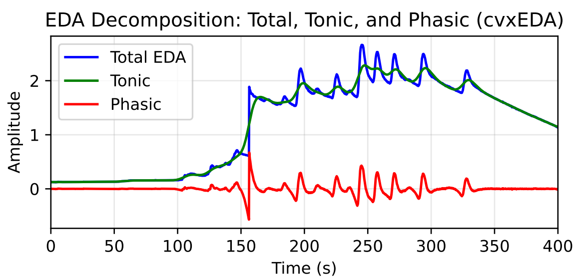
The structure of this paper is as follows: Section II reviews the fundamentals of EDA, compares related surveys, and discusses datasets. Section III details the methodology, including artifact removal, decomposition algorithms, and machine learning/deep learning models. Section IV presents a comprehensive benchmark performance comparison. Section VI discusses challenges and future directions. Finally, the contributions of the paper are summarized.

II. Background

A. EDA



(a) An illustration of EDA response



(b) An illustration of EDA decomposition

Fig. 1: EDA signal analysis. (a) shows a typical EDA response, and (b) shows the decomposition of an EDA signal.

EDA, also known as Galvanic Skin Response (GSR), is an indicator that reflects physiological states by measuring the dynamic changes in skin conductance. This signal primarily originates from the regulation of sweat gland activity by the SNS and is typically collected using electrodes on the palm or fingertips. As a sensitive physiological indicator, EDA can objectively reflect an individual's physiological responses during psychological processes such as cognitive effort, emotional arousal, and stress. As shown in Figure 1a, the EDA signal exhibits non-stationary characteristics, with a typical amplitude range of 2–20 µS and an inter-individual baseline difference of approximately 1–3 µS. A typical skin conductance response has a peak rise time of about 1–3 seconds and a half-recovery time of 2–10 seconds [15], [34].

As shown in Figure 1b, the EDA signal can be decomposed into two main components: a tonic component and a phasic component. The tonic component, also known as the Skin Conductance Level (SCL), represents the slow fluctuations of skin conductance over a longer time scale (frequency range 0.05–0.5 Hz). It primarily reflects the individual's overall, background level of arousal and is susceptible to long-term factors such as the environment or the individual's physiological baseline. In contrast, the phasic component manifests as a series of rapid, transient peak changes (frequency range 0.5–2 Hz), known as the Skin Conductance Response (SCR). SCRs are usually triggered by sudden SNS activation, for example, when an individual confronts a stressor. The superposition of multiple SCRs is the main reason for the complex morphology of the EDA signal; however, the amplitude and frequency of these responses are also directly related to the intensity of external stimuli. The reason EDA is particularly effective in stress detection is that the state of stress inherently triggers the SNS-dominated "fight-or-flight" response, leading to a significant increase in sweat gland secretion activity. This change can be clearly captured and quantified by the EDA signal. A more in-depth discussion of EDA decomposition algorithms will be provided in Section III-B.

B. Comparison with other surveys

As shown in Table I, we have systematically reviewed existing surveys closely related to the theme of this study and conducted an in-depth analysis from two dimensions: Contributions and Limitations. Our criteria for selecting surveys focused on two core types of literature: those that explicitly target EDA as the main research object, and those that concentrate on the stress detection scenario. The "Cited Articles" metric in the table reflects the breadth of literature research in each survey, while "Focus" clarifies its core area of attention.

Existing research can be broadly classified into two categories. The first category consists of **EDA-centric surveys**, such as [35], which pioneered the focus on EDA signal acquisition and processing techniques but did not extend to downstream modeling and applications. Subsequent works, while beginning to cover the complete technical pipeline, often had narrow application scenarios, such as focusing only on emotion recognition in educational contexts [37] or specific care environments [39], limiting their generalizability. Among these, [36] is the closest in focus to our study, but it has a fundamental problem in defining its research scope by conflating the two distinct fields of "stress detection" and "emotion recognition." The second category comprises **surveys centered on multimodal stress detection**. These works are generally more mature and systematic, tending to summarize and categorize existing work in the field from a macro perspective, either by process or by signal type [9], [41]. However, due to the vastness of the multimodal affective computing field itself, these surveys often struggle to provide an in-depth, detailed analysis of the specific modality of EDA, resulting in a lack of specificity.

C. Datasets

To promote reproducible research and enhance the generalization ability of models, a systematic review of existing public datasets is crucial. As shown in Table II, we have compiled detailed information on 21 public datasets that contain EDA signals and are relevant to stress detection. Although some previous works have summarized datasets in the field

Ref./Year	Cited Articles	Signal Modality	Core Focus	Main Contributions	Main Limitations
[35]/2020	89	EDA Only	EDA Data Acquisition and Signal Processing	1. Systematically reviews EDA data acquisition and signal processing techniques. 2. Summarizes recording devices, electrode technology, and advances in non-linear analysis. 3. Discusses data quality assessment and artifact processing methods.	1. Focuses on the signal processing front-end, without covering model construction. 2. Lacks a horizontal comparison and guiding recommendations for different methods (e.g., decomposition algorithms).
[36]/2020	103	EDA Only	EDA with ML for Stress Detection	1. First systematic review of the full pipeline for EDA-based stress detection. 2. Summarizes and compares the performance of existing features and models. 3. Emphasizes the balanced advantage of EDA in terms of accuracy and computational cost.	1. Methodological summary (e.g., datasets, decomposition algorithms) is not sufficiently in-depth. 2. Conflates research on emotion recognition with stress detection. 3. The study has not undergone peer review (preprint).
[37]/2021	86	EDA Only	EDA-based Emotion Recognition in Learning Contexts	1. Outlines the process of EDA-based emotion detection in educational contexts and points out the lack of standardization. 2. Systematically summarizes contradictory findings in studies linking EDA and learning.	1. The application scenario is limited to the educational field, lacking generalizability. 2. Methodological discussion is relatively basic, with an incomplete comparison of models and features.
[9]/2021	75	Multimodal Physiological Signals	Mental Stress Detection using Wearable Devices	1. Classifies by sensor type and application environment, summarizes the correlation between physiological signals and stress, and highlights the importance of EDA and HRV. 2. Reviews the strengths, limitations, and future directions of the field.	1. Does not systematically summarize public datasets. 2. Overly descriptive analysis, lacking quantitative comparisons and in-depth methodological discussion.
[38]/2023	87	Multimodal Physiological Signals	Generalizability of ML for Stress Detection	1. Follows a standard ML pipeline to systematically review commonly used datasets and pipelines. 2. Is the first to explicitly identify and emphasize the problem of poor generalization ability in the field, caused by models relying on single, small datasets.	1. The summary of datasets and methods is still mainly descriptive, lacking deep insights. 2. The analysis of the generalization problem remains superficial, without an in-depth exploration of solutions.
[39]/2025	74	EDA Only	EDA-based Stress Sensing in Care Environments	1. Bridges theory and practice, emphasizing the potential of EDA in special care settings (e.g., dementia). 2. Provides a detailed statistical analysis of subjects, devices, and other elements in the reviewed studies.	1. The reviewed studies are mostly on healthy individuals in laboratory settings, with insufficient coverage of real care contexts. 2. The analysis is primarily based on simple statistics, severely lacking in methodological summary and insight.
[40]/2025	185	Multi-source Data (Physiological, Behavioral, etc.)	Application of DL in Multimodal Stress Detection	1. Comprehensively overviews the application of DL in multimodal (physiological, speech, facial, social media, etc.) stress detection, highlighting its advantage in automatic feature extraction. 2. Details the methods, performance, and public datasets for each modality.	1. Is only a literature review, lacking an empirical benchmark comparison. 2. The focus on EDA is relatively limited.
[41]/2025	202	Multimodal (Primarily HR, EDA)	Continuous Stress Monitoring in Knowledge Work Environments	1. Reviews the integrated application of physiological, behavioral, and environmental data in knowledge work scenarios. 2. Emphasizes the importance of personalized models (e.g., federated learning) and user acceptance.	1. The scenario is limited to knowledge work, lacking cross-domain comparisons. 2. Discussions on topics like user acceptance are overly theoretical, lacking empirical data support.
This Study	203	**Primarily EDA**	**A Comprehensive Survey and Empirical Benchmark for EDA-based Stress Detection**	1. **First** to provide an end-to-end systematic survey and resource compilation for EDA-based stress detection, with an in-depth integration of 21 datasets. 2. **First** to establish a standardized cross-scenario evaluation pipeline, providing a benchmark for model performance and generalization ability through large-scale empirical comparisons.	1. The benchmark experiments do not cover more advanced models such as self-supervised learning or domain adaptation.

TABLE I: Comparison of Related Surveys on Stress Detection.

of affective computing [42]–[44], they do not fully meet the specific needs of current research in EDA-based stress detection. For example, [42] primarily focuses on cognitive load rather than the broader context of psychological stress; [43] has a limited scope, including only 9 datasets, of which only 5 involve EDA; and while [44] comprehensively summarizes 58 stress datasets, it does not specifically focus on research scenarios where EDA is the core modality. This study, by providing a detailed catalog, catalogs the key attributes of each dataset, including the number of subjects, experiment duration, included signal modalities, collection devices and frequencies, core experimental paradigms, and data access links, aiming to provide researchers with a comprehensive and convenient reference guide.

However, despite these datasets being publicly available, they still present a series of challenges in practical application. The first is **data accessibility barriers**. Some datasets (e.g., VeroBIO, EmpathicSchool, StressID, K-EmoPhone, ForDigitStress) are not available for direct download but require researchers to submit an application form or contact the original authors directly via email, which undoubtedly increases the complexity and time cost of data acquisition. The second is **technical usability barriers**. Some early datasets (e.g., RWDriving) are limited by the acquisition devices and experimental protocols of their time, which may not align with current technical standards. Other datasets (e.g., SWELL) use special data formats (S00 format) that require specific software or code libraries for parsing, adding an extra burden to data preprocessing. Finally, **the release status of some datasets is still unclear**. Several highly valuable datasets, such as CIRVR [45], which focuses on the job interview stress of autistic youth; LAUREATE [46], a longitudinal study dataset with high ecological validity; MMSD [47], a multimodal dataset with salivary cortisol as the gold standard; and MuSE [48], which incorporates real exam stressors, have been detailed in the literature with intentions to be open-sourced. However, at the time of this survey's completion, they have not yet been publicly released, and their future availability is uncertain. These factors collectively constitute the practical difficulties that researchers currently face when utilizing public datasets.

III. Methodology

In the early 2010s, a series of pioneering studies laid the methodological foundation for automated stress detection using wearable EDA sensors. The research by Setz et al. [70] demonstrated that an accuracy of up to 82.8% could be achieved in distinguishing psychological stress from cognitive load using only the EDA signal, providing the first strong evidence for the feasibility of EDA in the field of stress detection. [71] extended the research perspective to real-world data, exploring the feasibility of automatically identifying stress states in daily work scenarios. During this period, some of the core challenges in the field were identified and preliminarily explored. In terms of **multimodal fusion**, Choi et al. [72] combined EDA with other physiological signals (heart rate, respiration, EMG) and effectively distinguished different physiological states through unsupervised clustering. Kurniawan et al. [73], under user-independent conditions, confirmed that EDA exhibits superior performance and stronger generalization ability compared to speech signals. Regarding **personalized models**, to address the significant individual differences in physiological responses, [74] successfully constructed effective personalized adaptive models by improving the SVM loss function and selecting similar training samples based on electrodermal reactivity. Furthermore, due to the limitations of early hardware technology, a significant amount of research also focused on hardware innovations for data acquisition [75]–[77].

A. Artifact Removal

Motion Artifacts are non-physiological deviations introduced into the EDA signal by physical activity. Their sources include not only skin deformation under the electrodes but also muscle activity far from the measurement site, making them a major source of interference affecting signal quality. The methodology for handling motion artifacts primarily follows two technical paradigms: one is to directly correct the signal to suppress or **remove artifacts**, and the other is to segment the signal and perform quality assessment to **identify and discard** contaminated segments.

In the area of artifact removal, traditional non-adaptive methods, such as exponential smoothing [78] or low-pass filters [79], are simple but struggle to handle sudden artifacts with amplitudes much larger than normal EDA signals. They may also indiscriminately smooth artifact-free signal segments, causing information loss. To overcome these limitations, subsequent research has proposed various adaptive denoising methods. Chen et al. [80] proposed an adaptive denoising framework based on the stationary wavelet transform, which outperformed traditional methods but was sensitive to parameter tuning. Tronstad et al. [81] developed an extended Kalman filter based on a biophysical model, achieving effective suppression of noise and artifacts. Shukla et al. [82] modeled the coefficients in the wavelet domain, replacing the computationally expensive Gaussian Mixture Model (GMM) used in Chen's work with a Laplace distribution, achieving improvements in both artifact attenuation and signal fidelity. Additionally, for specific scenarios like driving, researchers have proposed strategies that dynamically weight and fuse EDA signals from both hands to effectively filter out sharp spike interference caused by actions such as turning the steering wheel [83], [84].

In the area of artifact identification, the research goal is to automatically locate and exclude invalid data segments. Taylor et al. [85] proposed a supervised learning method that extracts features from data segments and uses an SVM for classification to distinguish signals from artifacts. However,

Dataset Name	Year	Subjects	Duration	Signals	Device/Frequency	Description	Link
1. RWDriving [49]	2008	27	50–90 min	EKG, EMG, EDA, RESP	FlexComp/31 Hz	Real-world driving scenario	link
2. SWELL [50]	2014	25	138 min	EDA, HRV, ECG	Mobi device (TMSI)/2048 Hz	Knowledge work, stress induced by email interruptions	link
3. UTD [51]	2017	20	37 min	Acc, Temp, EDA, SpO2, HR	Affectiva/8 Hz	Physical, cognitive, and emotional stressors	link
4. WESAD [52]	2018	15	120 min	ACC, BVP, EDA, HR, IBI, TEMP	RespiBAN/700 Hz, Empatica E4/4 Hz	Baseline, stress, amusement, and meditation states	link
5. AffectiveROAD [53]	2018	13	120 min	ACC, BVP, EDA, HR, IBI, TEMP	Empatica E4/4 Hz	Real-world driving scenario	link
6. CLAS [54]	2019	62	30 min	ECG, PPG, EDA, 3D Accelerometer	Shimmer GSR Unit/256 Hz	Mathematical, cognitive, and emotional stressors	link
7. Toadstool [55]	2020	10	60 min	ACC, EDA, BVP, IBI, HR, TEMP	Empatica E4/4 Hz	Playing the game "Super Mario Bros."	link
8. MAUS [56]	2021	22	35 min	ECG, Fingertip-PPG, Wrist-PPG, GSR, IBI	Procomp Infiniti/256 Hz	N-back cognitive task	link
9. VeroBIO [57]	2022	55	NA	EDA, ECG, BVP, TEMP, ACC	Empatica E4/4 Hz	Public speaking anxiety	link
10. EmpathicSchool [58]	2022	20	81 min	EDA, HR, BVP, TEMP, IBI, ACC	Empatica E4/4 Hz	Completing various tasks in front of a computer	link
11. Wearable Exam Stress [59]	2022	10	4.5 hrs	EDA, HR, BVP, TEMP, IBI, ACC	Empatica E4/4 Hz	Real-world exam scenario	link
12. Nurse Stress Prediction [60]	2022	15	during 1 week	EDA, HR, TEMP, ACC, IBI, BVP	Empatica E4/4 Hz	Hospital work scenario	link
13. StressID [61]	2023	65	35 min	ECG, EDA, RESP	Bio SignalsPlux system/500 Hz	11 tasks including cognitive, emotional, and speech	link
14. UBFC-Phys [62]	2023	56	9 min	BVP, EDA, RPPG, PRV, TEMP	Empatica E4/4 Hz	Speech and mental arithmetic tasks	link
15. BIOSTRESS [63]	2023	28	90 min	EDA, ACC, TEMP, BVP, HR, IBI	Empatica E4/4 Hz	Stressful interview and final exam scenarios	link
16. K-EmoPhone [64]	2023	77	7 days	EDA, ACC, HR, TEMP, RRI and so on	Microsoft Band 2/5 Hz	Seven days of daily life monitoring	link
17. ForDigitStress [65]	2023	40	75 min	EDA, PPG, Voice	IOM biofeedback sensor/5 Hz	Job interview scenario	link
18. MUMTG [66]	2024	20	NA	EDA, BVP, HR, IBI, TEMP, Pupil Diameter	Empatica E4/4 Hz	Building blocks task	link
19. EmpaticaE4Stress [67]	2024	29	40 min	ACC, BVP, EDA, HR, IBI, TEMP	Empatica E4/4 Hz	Simulated work scenario	link
20. WorkStress3D [68]	2024	20	8.75 h	EDA, BVP, TEMP, ACC	Empatica E4 Wristband/4 Hz	In-office work day	link
21. EmoPairCompete [69]	2024	28	72 min	EDA, HR, BVP, TEMP, ACC	Empatica E4 Wristband/4 Hz	Competitive Tangram puzzle task	link

TABLE II: A Comprehensive Summary of 21 Publicly Available Datasets for Stress Detection Featuring EDA. Signal Abbreviations: Electrocardiogram (EKG/ECG), Electromyogram (EMG), Electrodermal Activity (EDA), Respiration (RESP), Heart Rate Variability (HRV), Accelerometer, Temperature (TEMP), Blood Oxygen Saturation (SpO2), Blood Volume Pulse (BVP), Heart Rate (HR), Inter-Beat Interval (IBI), Photoplethysmogram (PPG), Remote Photoplethysmography (RPPG), Pulse Rate Variability (PRV).

Zhang et al. [86] pointed out that supervised methods are highly dependent on manual annotation and are difficult to scale to large studies. Through a detailed comparison of various supervised and unsupervised learning methods, they demonstrated the effectiveness of unsupervised learning in artifact detection tasks. On the other hand, considering the "black-box" nature of machine learning methods, Kleckner et al. [87] proposed a transparent identification method based on four heuristic physiological rules, enhancing the algorithm's interpretability.

B. EDA Decomposition

Decomposing the EDA signal into its underlying tonic and phasic components is a key step in signal analysis, and a large number of decomposition algorithms have emerged. In early research, Bach et al. pioneered the use of a linear convolution model for the time-series analysis of rapid event-related SCRs, effectively resolving the issue of overlapping response peaks [88]. Subsequently, they enhanced the model's time-invariance by introducing physiological constraints [89] and used Dynamic Causal Modeling (DCM) to simulate sudomotor nerve activity in anticipatory responses [90]. Concurrently, Benedek et al. developed a technique based on non-negative deconvolution, ensuring the non-negativity of the driver signal through Gaussian elimination [91], [92]. These pioneering model-driven methods laid the foundation for subsequent research [93]–[95]. Later research gradually shifted towards frameworks based on sparse representation and optimization theory. Chaspari et al. introduced knowledge-driven dictionary learning and greedy recovery algorithms, improving the recovery accuracy of SCRs [96]. Bach et al. employed a matching pursuit algorithm to capture sparse, spontaneous SCR events to infer tonic sympathetic arousal levels [97]. Among these, the cvxEDA method, developed by Greco et al. and based on convex optimization, became a significant milestone [98]. This model explicitly models the EDA signal as a superposition of phasic, tonic, and noise components, and significantly improves the algorithm's robustness to noise through sparse recovery and error minimization strategies. The latest research advancements focus on more complex system identification models and near-real-time applications, encompassing time-frequency analysis techniques [99], compressed sensing [100], a non-negative sparse deconvolution algorithm based on an overcomplete dictionary (sparsEDA) [101], as well as Bayesian inference models based on state-space physiological representations [102]–[104] and a multichannel decomposition framework [105].

Although numerous decomposition algorithms have been proposed, only a few have been conveniently released as open-source software. Among them, Ledalab (including CDA and DDA), SCRalyze, and PsPM are primarily provided as standalone applications or toolboxes for specific software (like MATLAB), which to some extent limits their flexibility and applicability in cross-platform, automated analysis workflows [106]–[108]. In contrast, both cvxEDA and sparsEDA have been released as independent open-source code libraries on GitHub, greatly facilitating their widespread adoption and dissemination within the research community [109], [110].

Currently, some studies have compared the performance of different decomposition methods in practical scenarios [33], [111]–[113]. [112] found that in psychological stress detection scenarios, sparsEDA performed best in both computational efficiency and classification accuracy. [33] reported that in a Stroop test scenario, the performance of CDA was slightly better than cvxEDA. Meanwhile, [113], in an emotion recognition study involving six algorithms, found TVSymp to be the most effective. Although the optimal algorithm identified varies across studies, there is a consensus that the performance difference among mainstream algorithms (e.g., CDA, cvxEDA, sparsEDA) in the final classification task is typically within only 1-2 percentage points. This suggests that existing mainstream decomposition methods can all effectively decompose the EDA signal, and the differences in final application performance may stem more from the downstream feature extraction and modeling strategies rather than the superiority of the decomposition algorithms themselves.

Fortunately, the emergence of comprehensive physiological signal processing libraries such as NeuroKit2 [114] and BioSPPy [115] has greatly simplified the application of these advanced methods. As open-source Python toolkits, they integrate core decomposition functions like cvxEDA and sparsEDA, providing researchers with standardized, easy-to-call APIs, which significantly lowers the technical barrier for research.

C. Nonlinear Methods

In addition to traditional time-domain and frequency-domain feature extraction, a series of methods based on nonlinear dynamics theory have been proposed to capture the intrinsic correlation between the EDA signal and stress states at a deeper level. Early research primarily focused on frequency-domain analysis. POSADA-QUINTERO et al. [116] utilized power spectral density analysis to propose an index named EDASymp, defined as the normalized power of the EDA signal within the 0.045–0.25 Hz frequency band. Building on this, they further employed time-varying complex demodulation for time-frequency analysis to propose the more sensitive TVSymp index. By quantifying the average of the time-varying spectral amplitude in the 0.08–0.24 Hz band, it allows for more precise tracking of the dynamic changes in sympathetic tone [99]. Other studies have explored more concise nonlinear features. [117] proposed two computationally simple features within short-term windows—the slope of the linear regression of the EDA signal and the area between the signal and this regression line—and demonstrated that these two features can efficiently distinguish between different states of relaxation.

Recent research trends have shifted towards utilizing complexity science to quantify the dynamic properties of the EDA signal. [118] proposed a nonlinear analysis method called ComEDA, aimed at quantifying the dynamic spatial complexity of the EDA signal. The algorithm first maps the one-dimensional EDA time series into a high-dimensional dynamic space using phase space reconstruction techniques. It then employs angular distance to quantify the relationships between points on the attractor trajectory and finally measures the overall complexity of the system by calculating the quadratic Rényi entropy of this distance distribution. Experiments have shown that ComEDA exhibits superior performance in distinguishing between stress and resting states, as well as different levels of stress. As an extension, [119] further proposed a multi-scale version, MComEDA. This method first generates a series of sequences at different time scales from the original EDA signal through a coarse-graining process. It then applies the single-scale ComEDA algorithm to each scale, forming a multi-scale trend plot, and uses the area under the curve as the final MComEDA index. Research has confirmed that MComEDA outperforms both EDASymp and the original ComEDA in the task of distinguishing between different levels of emotional arousal and valence.

D. Unimodal

To deeply investigate the feasibility and current state-of-the-art of using only EDA for stress detection, we have systematically reviewed representative studies in this field, as detailed in Table III. The table provides a detailed account of the datasets, preprocessing pipelines, feature engineering, model selection, validation methods, and final performance for each study. The upper part focuses on ML methods, while the lower part covers emerging DL applications. The preprocessing pipelines are summarized in strict accordance with the described procedures.

Table III reveals several key trends and challenges in the field. In terms of **dataset selection**, there has been a clear shift from early self-collected data to public datasets like WESAD, which has greatly promoted the reproducibility of results. However, regarding **methodology**, preprocessing and feature engineering still lack recognized standards, making cross-study performance comparisons challenging. In terms of models, Support Vector Machines (SVM) and ensemble learning models (such as XGBoost) have shown strong competitiveness among ML methods. Regarding **evaluation standards**, considering the strong individual variability of physiological signals, Leave-One-Subject-Out (LOSO) cross-validation or strict user-independent testing is gradually becoming the gold standard for evaluating the true generalization ability of models.

Among the many ML studies, some works are particularly noteworthy for their unique approaches. [124], to address the issue of individual differences, proposed a "Localized Supervised Learning" framework. By clustering data, it trains specialized classifiers for different physiological response patterns, achieving performance significantly superior to that of global models. It is worth noting that the optimal model often varies across different subgroups. [125] systematically compared Deep Support Vector Machines (DSVM) with various traditional classifiers, confirming the superiority of DSVM. It also explored the impact of the analysis window length, pointing out that model performance tends to saturate after 5 seconds, and found that whether the windows overlap has no significant effect on the final performance. Notably, several studies have consistently shown that the performance of unimodal EDA is outstanding. The research by [16] and [17] both demonstrated that on multiple public datasets, the classification performance using only the EDA signal was even better than that of other single modalities or multimodal combinations. Furthermore, [130] introduced the concept of "Time to Detection" (TTD), which is more relevant to practical applications. It refers to the time from the occurrence of a stress event to its detection, defined as $\mathrm{TTD} = t_{\text{length}} + (\omega - 1)t_{\text{step}}$ (where t_{length} is the sliding window length, t_{step} is the sliding step size, and ω is the number of consecutive data segments required to make a final decision). This aims to provide a basis for balancing speed and accuracy in the design of real-time warning systems.

In recent years, end-to-end deep learning methods have also begun to be applied to unimodal EDA stress detection, aiming to reduce the reliance on manual feature engineering. [132], addressing the limited size of public datasets, developed an EDA biophysical model to generate high-quality synthetic data by simulating signal characteristics under different stress levels, effectively augmenting the training set. [133] utilized Neural Architecture Search (NAS) technology to automatically design and optimize a lightweight convolutional neural network named TinyStressNet. This model significantly reduces computational complexity and model size while maintaining performance comparable to that of large benchmark models, demonstrating great potential for real-time stress monitoring on resource-constrained edge devices.

Although this survey focuses on stress detection, some research from the field of EDA-based emotion recognition also offers valuable insights. [34] summarized 25 previous studies and systematically compared the importance of 40 EDA features in emotion recognition tasks, finding that features from the acoustic domain, such as Mel-Frequency Cepstral Coefficients (MFCCs), performed excellently. This opens up new avenues for feature engineering in stress detection. [134] was the first to introduce Graph Signal Processing (GSP) techniques to EDA analysis, transforming the traditional time-series signal into a graph structure. This

PaperDataset(s)	Preprocessing	Features (Domain, Selection Method, Count)	Model(s)	Validation	Reported Performance
[120] Self-collected; 5 mental arithmetic tasks	50Hz notch, min-max norm, 500ms moving avg, 3.5s window	55 (Time, Freq, Time-Freq, Quefrency) (PCA, 90% variance)	**Combined SVM**, PNN	Double CV, 5-Fold	90.0%/80.2% Acc (3/5-class)
[121] Self-collected; running & TSST	5Hz low-pass filter, detrend, 3 min window	33 (Time) (Wrapper method, 5)	**LDA**, QDA, SVM, KNN	LOOCV	95.1% Acc (Stress type classification)
[122] RWDriving	Low-pass filter, 5 min window	18 (Time) (Fisher projection, 2)	LDA	LOOCV	81.82% Acc (3-class stress level)
[123] WESAD	4Hz downsample, 2Hz low-pass, 60s window (30s overlap)	195 (Multi-domain) (Pearson corr ¿ 0.9, 9)	**XGBoost**, DT, RF, AdaBoost	LOSO	92.38%(chest)/89.92%(wrist) F1
[124] Self-collected; pre-surgery patients	5Hz low-pass filter, detrend, 5s window	57 (Time, Wavelet) (Stat. test + wrapper, 4-7)	Localized Supervised Learning	30/11 subject split	85.06% Acc (3-class stress level)
[125] Self-collected; watching stressful movies	DDA (Ledalab) decomp, 1-40s window	23 (Time, Morphological, Freq) (None)	**DSVM**, various traditional classifiers	70/15/15 split	92% F1 (2-class)
[126] Self-collected; academic stress	z-score norm, 5 min sample	Raw signal input (LDA for dimensionality reduction)	**SVM**, LDA, k-NN	5-fold CV	100% Acc (GSR), 96% Acc (e-nose)
[33] Self-collected; Stroop test	Downsample, smooth, z-score norm, CDA/cvxEDA decomp.	4 (Time) (None)	ELM	6-fold CV	95.56%(CDA)/94.45%(cvxEDA) Acc
[127] WESAD, VERBIO	30s window (no overlap), NeuroKit decomp.	7 (Time) (None)	KNN, LR, RF	90/10 subject split	85.3%(VerBIO)/85.7%(WESAD) Acc
[16] WESAD, CLAS	30s window (no overlap), cvxEDA decomp.	7 (Time) (None)	**SEL**, SVM, KNN, RF, NB, LR	Not Specified	86.4%(WESAD)/72.3%(CLAS) Acc
[128] WESAD	8Hz upsample, 60s window, 0.25s stride, cvxEDA decomp.	87 (Time) (Forward selection, 5)	**AdaBoost**, RF, SVM	LOSO	97.03% Acc
[129] Self-collected; TSST	cvxEDA decomp.	11 (Time, Freq) (SVM-RFE, 8 for 2-class, 6 for 4-class)	**SVM-RFE**, RF, LDA	LOSO	94.62%(2-class)/75.00%(4-class) Acc
[130] WESAD	12s window (5s overlap), norm	$3(3m+1)$ (Freq) (None)	DT, 1/10-NN, RF, SVM, Bagged SVM, AdaBoost	LOSO	84.87%(2-class)/69.89%(3-class) Acc
[14] CLAS, UTD, VERBIO, WESAD	30s window (no overlap), cvxEDA decomp.	7 (Time) (None)	KNN, SVM, NB, LR, RF	90/10 subject split	92.9%(VerBIO)/86.5%(WESAD) Acc
[131] 4 public, 2 private datasets	Manual artifact rejection, median filter, z-score, 30s window	47 (Multi-domain) (MIC, 40)	**Voting Ensemble** (GBT, RF, AdaBoost)	Train on combined, test on unseen	89.1%(Combined)/84.6%(Unseen) Acc
[17] 6 public datasets	Various (windowing, filtering, cvxEDA decomp.)	7 (Time) (None)	FNN, CNN, Wide and Deep Model	90/10 subject split	¿88% Acc across all datasets
[132] Self-collected; physical activity	2min window (1min overlap)	End-to-End	FCN	Train on synthetic, test on real	96% (Synthetic) / 84% (Real) Acc
[133] WESAD, T-Test, AffectiveROAD, DS-3	4Hz resample, band-pass filter, 30s window	End-to-End	**NAS** for TinyML	90/10 subject split	85.98% Acc (on combined dataset)

TABLE III: A Systematic Summary of Unimodal Studies Utilizing Only EDA for Stress Detection. The best-performing model as reported in the original paper is indicated in **bold**; entries are not bolded if multiple models were reported as optimal under different conditions. **Abbreviations**: CNN (Convolutional Neural Network), CV (Cross-Validation), DT (Decision Tree), ELM (Extreme Learning Machine), FCN (Fully Convolutional Network), FNN (Feedforward Neural Network), GBT (Gradient Boosted Trees), KNN (k-Nearest Neighbors), LDA (Linear Discriminant Analysis), LOSO (Leave-One-Subject-Out), LOOCV (Leave-One-Out Cross-Validation), LR (Logistic Regression), MIC (Maximal Information Coefficient), NAS (Neural Architecture Search), NB (Naïve Bayes), RF (Random Forest), SEL (Stacking Ensemble Learning), SVM (Support Vector Machine), TSST (Trier Social Stress Test).

work demonstrated the superiority of GSP in capturing complex nonlinear patterns in the EDA signal, providing a new theoretical perspective for future model innovation.

E. Insights from Multimodal Methods

Although the core of this survey is to demonstrate the high efficiency and feasibility of unimodal EDA in stress detection, a systematic review and analysis of multimodal research can provide valuable insights and inspiration for the EDA field. By integrating multiple physiological signals, multimodal methods aim to construct a more comprehensive representation of physiological states. Their advancements in sophisticated modeling techniques, complex experimental paradigm design, and considerations for practical applications offer valuable ideas for deepening EDA-related research.

Multimodal stress detection frameworks typically fuse EDA with other physiological indicators to capture a richer spectrum of stress responses. These auxiliary signals mainly include: Heart Rate (HR) and Heart Rate Variability (HRV), obtained via electrocardiogram (ECG) [83], [135]–[137] or photoplethysmography (PPG) [138]–[145], to quantify the balance of the autonomic nervous system (ANS); Skin Temperature (ST) [138]–[140], [142], which reflects peripheral vasoconstriction caused by SNS activation; electromyography (EMG) [136], [143] to quantify muscle tension; respiration signals (RSP) [136], [143], [146] to capture changes in breathing rhythm; and electroencephalography (EEG) [136], [146], which directly reflects central nervous system (CNS) activity. Additionally, the accelerometer signal serves as crucial contextual information, essential for distinguishing between similar physiological arousals caused by physical activity and psychological stress [142].

In the traditional ML paradigm, multimodal research typically follows a "feature extraction-fusion-classification" pipeline. Features are extracted from each modality and then jointly fed into a machine learning classification model. Although some studies involve special nonlinear feature calculations, the fundamental ML pipeline remains unchanged [142]. Furthermore, the processing of EDA is often relatively crude, with only a few studies separating its components [139]–[141], [147]. However, the introduction of DL has brought about profound changes in feature representation and model construction. As early as 2013, studies were already comparing the effectiveness of different neural networks [148]. Some research feeds traditionally hand-crafted features into neural networks like CNNs, RNNs, or LSTMs [145], [146], [149], [150]. From a methodological standpoint, this strategy fails to fully leverage the core advantages of DL. The strength of the ML paradigm lies in the strong interpretability of its models and features, but its performance is highly dependent on the quality and complexity of feature engineering. Conversely, the core advantage of the DL paradigm is its end-to-end learning capability, which allows it to automatically learn and discover hierarchical feature representations from raw signals, thus avoiding tedious and expert-knowledge-dependent manual feature extraction. The aforementioned hybrid paradigm, to some extent, combines the disadvantages of both approaches: it inherits the reliance on feature engineering from ML while introducing the interpretability challenges caused by the "black-box" nature of DL models. Although this method may be effective in specific applications, it is not the optimal solution. In contrast, some studies have adhered to the end-to-end philosophy of DL, directly feeding raw or lightly preprocessed physiological signals into deep networks, and have successfully demonstrated the feasibility and superiority of automatic feature learning directly from physiological time-series data [151], [152].

Although many studies have employed relatively basic network architectures like CNNs, RNNs, or LSTMs and have handled multimodal data through simple feature concatenation or channel stacking, the field has still seen the emergence of a number of innovative works in terms of model architecture and fusion strategies. At the feature fusion level, researchers are no longer satisfied with simple late-stage decision fusion but are exploring deeper feature interaction mechanisms. K et al. [153] systematically compared the effects of fusing ECG and EDA signals at different network depths (fully connected layers, mid-level convolution, deep convolution). Their empirical results showed that early feature interaction at the network's convolutional layers is more effective than fusion near the decision layer. Based on this insight, they further designed a "Multimodal Transfer Module" (MMTM) to perform cross-modal feature interaction and calibration in the higher-level feature parts of the network, combined with transfer learning to enhance model generalization performance [154]. Similarly, [155] designed parallel CNNs for ECG and EDA and constructed more information-rich "hierarchical feature sets" for each modality by concatenating features at low, middle, and high levels, then used MMTM for deep fusion at the intermediate level. Furthermore, hierarchical architectures have also been proven to be an effective fusion strategy. [156] designed a three-level hierarchical DNN that performs unimodal feature extraction, merging of information from similar sensors, and cross-sensor feature fusion at the shallow, middle, and deep layers, respectively, achieving excellent performance. Some research proceeds from the commonalities and specificities of modalities. [157] proposed a dual-channel framework where a "modality-invariant" channel is used to extract common representations of ECG and EDA, while a "modality-specific" channel uses a 1D-CNN to capture the unique discriminative features of each modality, finally fusing the complementary information at the decision layer.

In terms of network architecture innovation, researchers have drawn inspiration from advanced ideas in fields like computer vision. [158] took inspiration from the ResNeXt architecture, building a powerful end-to-end learning model by using multiple convolutional kernels of different sizes in parallel to capture multi-scale temporal features in the signals. To combine the advantages of CNNs in extracting local features with the ability of Transformers to model global dependencies, [159] proposed a CNN-Transformer hybrid model. [160] adopted a similar idea, using parallel Fully Convolutional Network (FCN) stacks to extract local spatial features while utilizing LSTM units to handle temporal dependencies in the high-level features. Additionally, unsupervised learning has been used to optimize the feature extraction process. [19] employed an autoencoder based on a pseudo-inverse learning algorithm to automatically learn features from raw physiological signals in an unsupervised manner. These features were then fed into an AdaBoost-based ensemble self-classification module for accurate identification. [161] transformed signals to the frequency domain and used multiple encoding layers of an autoencoder to extract and fuse features to construct a hierarchical frequency-domain representation, which was then input into a Convolutional Recurrent Neural Network (CRNN) with a Squeeze-and-Excitation (SE) module for classification.

Since models like CNNs were originally designed for processing two-dimensional image data, applying them directly to one-dimensional physiological signals may not fully leverage their powerful spatial feature extraction capabilities. Therefore, how to effectively convert one-dimensional time-series signals into two-dimensional image representations to better utilize advanced 2D-CNN architectures has become a promising research direction. One strategy is to use the recurrence properties of the signal. [162] converted EDA signals from the hand and foot, as well as heart rate data, into two-dimensional Continuous Recurrence Plots, and then fed these three-channel images into parallel CNNs for feature extraction and classification. Another mainstream approach is based on time-frequency analysis or mathematical transformations. [163] sequentially applied Fast Fourier Transform (FFT), a cube root transform, and a Constant Q Transform (CQT) to the signals to obtain feature maps, which were then stacked and fed into a CNN. Recently, encoding methods based on Gramian Angular Fields have received attention. [164] encoded multimodal signals into multi-channel two-dimensional images using the Gramian Angular Summation Field (GASF) and Gramian Angular Difference Field (GADF), which were then input into a 2D-CNN. This study confirmed that compared to the Markov Transition Field (MTF), GASF/GADF has higher computational efficiency. [165] also utilized these three techniques—GASF, GADF, and MTF—to convert one-dimensional feature vectors extracted from physiological signals into three separate two-dimensional matrices, which were stacked into a three-channel image and fed into a 2D-CNN, achieving superior classification results compared to traditional SVM methods. These signal-to-image conversion methods essentially encode the temporal dependencies, frequency characteristics, or phase information of time-series signals into the pixel-space structure of images, opening up new avenues for analyzing physiological signals using powerful vision models.

Beyond simple multimodal fusion, understanding the intrinsic correlation and substitutability between different physiological signal modalities is also of great significance. [166] conducted pioneering explorations in this area. By training a separate Transformer model for each modality and investigating its cross-modal generalization capability, they found strong generalization potential among physiological signals based on the ANS (e.g., ECG, EDA, RESP), whereas they were clearly distinct from motion signals (accelerometers). This provides profound physiological insights for sensor selection and model design.

In addition to the continuous evolution of model architectures and signal representations, researchers are increasingly focusing on the series of challenges faced when moving stress detection models from controlled laboratory settings to the complex real world ("in the wild"). A core problem is the inherent ambiguity of physiological signals, where physiological arousal caused by psychological stress and physical activity is often difficult to distinguish at the signal level. To solve this problem, context-based stress detection frameworks have emerged [138], [139]. The core idea is to use auxiliary sensors like accelerometers to identify the user's physical activity state and incorporate this contextual information into the model to help it correctly attribute physiological changes. This methodology is essentially similar to the strategy of identifying motion artifacts and removing the corresponding data segments [141]. However, when introducing contextual information, one must be wary of inappropriate designs, such as directly using the experimental phase as a feature [167] or relying on synchronized subjective reports [168]. Although these practices can improve metrics, they lack scientific rigor and practical value due to information leakage or impracticality. A more sophisticated solution is to design an adaptive framework that uses accelerometers or EMG signals as a "gating model" to dynamically activate the most suitable classifier branch according to the context [169]. Closely related to the complexity of context are the significant differences between individuals, which makes a single, generic model difficult to adapt to all users. Therefore, personalized modeling has become key to improving model robustness. The implementation paths are diverse, including adopting multi-task learning (MT-NN) at the model level to customize independent network branches for each user [170], combined with domain adaptation techniques to improve cross-subject generalization [171]; at the feature level, demographic information such as gender and age can be incorporated as auxiliary inputs [136], [150]; or at the data level, by clustering users and training specialized models within clusters to capture the response patterns of similar populations [143].

The journey of deploying models in the real world goes far beyond addressing context and individual differences. Researchers are beginning to systematically evaluate the generalization gap of models from the laboratory to real environments [144], [172], their generalization ability across different activities and populations [173], and even the critical yet often overlooked issue of reproducibility in the entire field [174]. These studies not only quantify the severity of the challenges but also bring profound insights. For example, a model trained on high-quality laboratory data may generalize better to noisy real-world data than a model trained directly on real-world data [172]. At the same time, to ensure the ecological validity of the research, the choice of research scenarios and stressors is becoming increasingly diverse and close to reality, covering various contexts from programming competitions [141] and public speaking [175] to speech activities [176] and simulated air raid sirens [177]. It has also begun to focus on specific populations (such as patients with mild cognitive impairment [137]) or specific types of stress (such as mental and physical stress [178]). A recurring and noteworthy point is that EDA has once again been proven to completely outperform multimodal approaches in some scenarios, for both laboratory and real-world data [18], along with its lower cost and convenience compared to multimodal setups [19]. This again corroborates the necessity of research on unimodal EDA.

In addressing data-level challenges, the problem of data scarcity severely constrains the effective training of complex models, as physiological signal datasets are generally small in scale and limited to single scenarios. To this end, creating a larger, more diverse training set by merging multiple existing datasets and training an ensemble model on it has been proven to be an effective strategy for improving generalization to unseen datasets [179]. Furthermore, borrowing from the successful experiences of other DL fields, the paradigm of ”pre-training” on a large-scale general dataset followed by ”fine-tuning” on a specific target task has been proposed, offering a promising approach to solve the data dilemma in this field [180]. In addition to the choice of models and data, the definition and quantification of ”stress” as a target are also evolving. Treating stress as a binary or multi-class classification problem is a simplification of a continuous psychological state. Consequently, an increasing number of studies are shifting towards regression tasks, aiming to predict a continuous stress score [181]. This approach not only allows for a finer-grained characterization of stress levels, preserving more individual information [182], but also provides a theoretical basis for implementing timely, graded early interventions. Finally, underlying all these high-level designs is the fundamental and critical methodological issue of signal windowing. The length and overlap rate of windows directly determine the stability of features and the temporal resolution of the model. Although it is generally believed that longer windows can improve accuracy by containing more information [149], this improvement has a saturation point [163], and it comes at the cost of real-time performance. Therefore, the selection of window parameters must be a deliberate trade-off between performance, real-time capability, and the amount of available data, based on the specific application requirements, and should be clearly justified as an important part of the model design.

IV. Experiments

To compensate for the general lack of empirical benchmarks in existing surveys and to systematically investigate the performance of current mainstream models in handling complex scenarios, this section constructs a standardized experimental pipeline. We aim to thoroughly evaluate the effectiveness and generalization capabilities of EDA-based stress detection models through two core experiments: 1) a Multi-Domain Supervised Experiment, designed to test the robustness of models in handling diverse and complex scenarios after merging multiple heterogeneous datasets; and 2) a Cross-Domain Generalization Experiment, aimed at evaluating the transfer performance of a model trained on one or more source domains to a completely unseen target domain. Together, these two experiments provide key empirical evidence for future model selection and generalization strategies in this field.

A. Experimental Pipeline

To ensure the reproducibility of the experiments and consistency in processing across different datasets, we designed and followed a concise yet efficient preprocessing pipeline. First, we did not use any explicit filters. This decision was based on the following considerations: first, as previously mentioned, modern decomposition algorithms like cvxEDA have inherent noise suppression capabilities sufficient to handle artifacts such as sudden signal spikes; second, considering the low sampling rate of devices widely used in this study, such as the Empatica E4 (e.g., 4Hz), according to the Nyquist sampling theorem, only low-pass filters below 2Hz have clear significance. However, the phasic component of the EDA signal, which is highly correlated with stress, can have frequencies up to 2Hz, and excessive filtering risks removing critical information; third, a streamlined pipeline helps establish a more easily reproducible benchmark. Therefore, our standardized pipeline is as follows: 1) **Resampling**: All EDA signals from all datasets were uniformly resampled to 4Hz. 2) **Normalization**: Z-score normalization was independently applied to the signal of each subject to reduce the effects of individual physiological baselines and inter-device differences. 3) **Decomposition**: The cvxEDA algorithm was used to decompose the raw EDA signal into its two core components, tonic and phasic. 4) **Windowing**: The decomposed signals were segmented into 30-second windows. This step was performed after decomposition to preserve the complete morphology of transient stress responses.

TABLE IV: Extracted EDA Features.

Type	Feature Group	Parameter	Description
Time Domain	Event Features	meanPeak	Average value of peaks
		avgRiseTime	Average rise time
		avgFallTime	Average fall time
		numPeaks	Number of peaks
		avgAbsoluteArea	Average area under the absolute curve
		numSignificantDrivers	Number of significant driver points
		meanDriver	Average value of significant driver signals
	Time Features	basicStatsTonic	Mean, std, rms, skew, kurt of the tonic component
		basicStatsPhasic	Mean, std, rms, skew, kurt of the phasic component
		basicStatsEDA	Mean, std, rms, skew, kurt of the EDA component
	Hjorth Features	HjorthTonic	Activity, Mobility, Complexity of the tonic component
		HjorthPhasic	Activity, Mobility, Complexity of the phasic component
		HjorthEDA	Activity, Mobility, Complexity of the EDA component
Frequency Domain	Frequency Features	basicStatsTonicFreqFeatures	Bandwidth, Centroid, Dominant, Min, Max Frequency of the tonic component
		basicStatsPhasicFreqFeatures	Bandwidth, Centroid, Dominant, Min, Max Frequency of the phasic component
		basicStatsEDAFreqFeatures	Bandwidth, Centroid, Dominant, Min, Max Frequency of the EDA component
	Frequency Amplitude Features	basicStatsTonicFreqAmplitude	Mean, std, rms, skew, kurt of the tonic component frequency amplitude
		basicStatsPhasicFreqAmplitude	Mean, std, rms, skew, kurt of the phasic component frequency amplitude
		basicStatsEDAFreqAmplitude	Mean, std, rms, skew, kurt of the EDA component frequency amplitude
	PSD Features	basicStatsTonicPSD	Mean, std, rms, skew, kurt of the tonic component PSD
		basicStatsPhasicPSD	Mean, std, rms, skew, kurt of the phasic component PSD
		basicStatsEDAPSD	Mean, std, rms, skew, kurt of the EDA component PSD
	Power Features	TonicPower	Average power of the tonic component
		PhasicPower	Average power of the phasic component
		EDAPower	Average power of the EDA component
Time-Frequency Domain	Wavelet Features	waveletTonic	Mean, std, rms, entropy, energy from all 5 sets of coefficients (cA4, cD4-cD1)
		waveletPhasic	Mean, std, rms, entropy, energy from all 5 sets of coefficients (cA4, cD4-cD1)
		waveletEDA	Mean, std, rms, entropy, energy from all 5 sets of coefficients (cA4, cD4-cD1)
Entropy Domain	Entropy Features	basicTonicEntropy	Sample, Permutation, Fuzzy, Spectral, Multiscale Entropy of the tonic component
		basicPhasicEntropy	Sample, Permutation, Fuzzy, Spectral, Multiscale Entropy of the phasic component
		basicEDAEntropy	Sample, Permutation, Fuzzy, Spectral, Multiscale Entropy of the EDA component

This study selected five public datasets: UTD, WESAD, MAUS, AffectiveROAD, and EmpaticaE4Stress. The first four datasets form the core of our experiments due to their diverse scenarios (covering cognitive, physical, and emotional stress, as well as real-world driving) and their widespread use. EmpaticaE4Stress, a new dataset simulating a real work scenario, was used as an independent target domain to evaluate the generalization ability of the models. Table V summarizes the statistical information of each dataset after windowing. All experiments strictly adhered to the ”subject-independent” principle, meaning all data segments from the same subject were strictly assigned to either the training, validation, or test set, to ensure that the model evaluation assesses its generalization ability to new users. In the multi-domain fusion experiment, we combined the first four datasets and partitioned the total sample size into training, validation, and test sets in an 8:1:1 ratio based on subject IDs.

TABLE V: Summary Statistics of Datasets Used in the Benchmark Experiments.

Dataset	Subject Split (Train/Val/Test)	Stress Segments	Non-Stress Segments
UTD	16/2/2	702	789
WESAD	9/2/2	326	1,147
MAUS	16/3/3	880	440
AffectiveROAD	9/2/2	1,823	1,494
Combined Dataset	50/9/9	3,731	3,870
EmpaticaE4Stress	23/6 (Train/Test)	1,396	754

At the model construction level, we designed two distinct technical paths: ML and DL. In the **ML Pipeline**, we first extracted a comprehensive feature set from each window, covering the time, frequency, time-frequency, and entropy domains (see Table IV). Subsequently, we used XGBoost to rank the features by importance and determined the optimal feature subset for each model through Top-K experiments. The **DL Pipeline** adopted an end-to-end strategy, using the raw EDA, tonic, and phasic sequences as a three-channel input. Each channel was then individually z-score normalized again to eliminate the absolute magnitude differences between the tonic and phasic components, allowing the model to autonomously learn feature representations and thus avoiding manual feature engineering.

B. Multi-Domain Fusion Analysis

Figure 2 shows the trend of AUC for each ML model on the Combined Dataset as the number of features increases. A significant finding is that, with the exception of SVM and DT, the performance of most models tends to saturate when the number of features reaches approximately 22. Adding more features beyond this point did not lead to a significant performance improvement, and for some models (such as MLP, RF, KNN), a slight performance degradation was even observed. This phenomenon reveals that for a 30-second EDA signal segment, a relatively compact feature subset (about 22 features) may be sufficient to capture the core information related to the stress state. This has important guiding implications for deploying efficient models on resource-constrained wearable devices.

Table VI and Table VII present the optimal performance of the ML and DL models on the Combined Dataset, re-

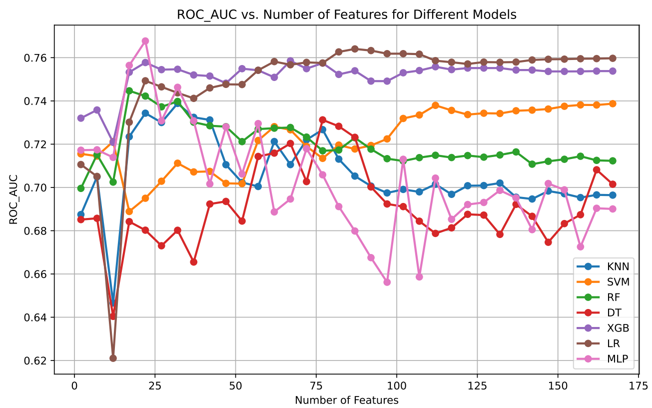


Fig. 2: ROC AUC vs. Number of Features for Different Models on the Combined Dataset Test Set.

TABLE VI: PERFORMANCE OF MACHINE LEARNING CLASSIFIERS ON THE COMBINED DATASET

Classifier (Features)	ACC (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUC (%)
KNN (32)	64.62	51.48	75.69	39.01	73.89
SVM (113)	66.10	52.55	80.49	39.01	73.88
RF (24)	66.10	53.73	78.28	40.90	74.49
DT (76)	65.64	55.06	74.30	43.74	73.35
XGB (75)	65.87	52.23	80.00	38.77	75.99
LR (87)	66.21	52.33	81.50	38.53	76.40
MLP (22)	**66.55**	**54.77**	**78.41**	**42.08**	**76.77**

spectively. The experimental results initially reveal two core findings. First, in terms of accuracy, the best-performing DL model (CNN, 72.35%) significantly outperformed all ML models (MLP, 66.55%). This directly reflects the fundamental difference between the two methodologies when integrating data from diverse scenarios. Traditional ML methods rely on a fixed, manually designed feature set. However, after fusing heterogeneous data from driving, cognitive tasks, social stress, etc., no single fixed set of features can be optimal for all scenarios simultaneously. In contrast, the end-to-end CNN model, with its automatic and hierarchical feature learning capabilities, can autonomously discover commonalities and specificities across scenarios from the raw signals, thus demonstrating stronger robustness when processing highly heterogeneous data.

Second, a more noteworthy phenomenon is that almost all models exhibited low sensitivity and high specificity. This indicates that the models are relatively good at identifying non-stress states but have significant missed detections for true stress states. The root of this phenomenon may not only be that non-stress segments are dominant in number or duration in the dataset but could be due to a deeper reason: the physiological signals of the "stress" state itself are highly heterogeneous. Different types of stressors (cognitive, emotional, physical) trigger SNS response patterns that vary greatly in detail. This makes the "stress" class in the feature space resemble a loose, vaguely bordered "nebula." In contrast, the "non-stress" state usually corresponds to a relatively stable physiological baseline, with more homogeneous signal patterns, forming a compact "star cluster" in the feature space. This undoubtedly increases the difficulty for a classifier to learn a decision boundary that can effectively encompass all stress samples, highlighting a challenge that future algorithm designs must focus on addressing.

TABLE VII: PERFORMANCE OF DEEP LEARNING MODELS ON THE COMBINED DATASET

Model	ACC (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUC (%)
MLP	69.17	63.59	74.34	66.50	76.57
CNN	**72.35**	**56.97**	**86.62**	**66.48**	**80.39**
ResNet18	67.92	43.50	90.57	56.62	79.02
RNN	67.01	49.17	83.55	58.92	72.05
GRU	72.01	58.87	84.21	66.94	78.52
LSTM	69.62	55.56	82.68	63.77	78.61
BiLSTM	70.19	56.97	82.46	64.78	75.81
ViT	70.42	56.97	82.89	64.96	78.41

C. Cross-Domain Generalization Analysis

For a more rigorous assessment of generalization ability, we selected the best-performing ML model from the previous stage (MLP with the 22 optimal features, as shown in Table VIII) and DL model (CNN). These models were trained on different source datasets and tested on the completely independent EmpaticaE4Stress (E4) dataset.

TABLE VIII: Top 22 Most Important EDA Features (Grouped by Type): 14 EDA, 7 phasic, 1 tonic.

Type	Features
Time Domain (5)	mean of phasic, standard deviation of EDA, mean amplitude of peaks, number of peaks, Hjorth activity of phasic
Frequency Domain (7)	mean of EDA PSD, standard deviation of EDA PSD, RMS of EDA PSD, RMS of EDA FFT amplitude, standard deviation of EDA FFT amplitude, kurtosis of EDA FFT amplitude, dominant frequency of phasic FFT
Time-Frequency Domain(8)	mean of phasic cA4, RMS of EDA cD3, standard deviation of EDA cD4, RMS of EDA cD4, energy of phasic cD3, entropy of tonic cA4, RMS of EDA cA4, RMS of phasic cD3
Entropy Domain (2)	Sample Entropy of phasic, Multiscale Entropy of EDA

As shown in Table IX and Table X, the results clearly reveal the severity of the domain shift problem. Almost all models trained on a single source domain showed a significant performance degradation when transferred to the completely independent EmpaticaE4Stress (E4) dataset. This directly confirms and explains why simply aggregating data in the multi-domain fusion experiment failed to achieve ideal results—the physiological patterns learned by the model are highly coupled with specific scenarios and experimental paradigms.

However, the most insightful finding from this analysis is that the model trained on the AffectiveROAD dataset exhibited the best transfer performance. The CNN model (AUC 72.42%) not only far surpassed the corresponding MLP model (AUC 67.38%) but even exceeded the baseline model trained on the target domain E4 (AUC 70.16%). This key result provides two profound insights for the future development of the field. First, **the ecological validity of the data is more critical than the sheer volume of data**. AffectiveROAD originates from a real-world driving environment, making its physiological response patterns more generalizable than those from highly controlled laboratory settings. This suggests that future research should shift its focus from blindly pursuing larger mixed datasets to identifying and utilizing core datasets with high ecological validity for pre-training and transfer. Second, **the knowledge learned by DL models is more transferable**. The successful transfer and outperformance of the baseline by the CNN strongly indicate that it learned a more abstract representation, closer to the ”essence of the physiological stress response,” from the high-quality source domain data. The ”purity” and generalizability of this representation are even higher than the knowledge learned directly from the target domain, which has a relatively singular scenario. In contrast, the capability of the MLP was limited by the initially selected 22 features, making it unable to ”tailor” a feature set most suitable for transfer for high-value data.

In summary, our benchmark tests systematically quantify the significant advantages of DL methods, even on unimodal EDA signals. However, a review of existing research (as in Table III) shows that end-to-end deep learning research specifically for unimodal EDA stress detection is still in its nascent stages. The contribution of this study is that, for the first time, through large-scale empirical comparison, it provides strong evidentiary support for shifting the research paradigm in this field from the traditional ”feature engineering + ML” to ”end-to-end DL.” It also offers a solid performance benchmark for future innovations in model architecture and domain adaptation research.

TABLE IX: PERFORMANCE OF MLP (22 FEATURES) TRAINED ON DIFFERENT DATASETS AND TESTED ON E4

Training Dataset	ACC (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUC (%)
Combined	61.66	68.04	74.29	62.76	66.34
UTD	63.68	74.04	69.16	79.66	61.77
WESAD	51.57	52.84	72.02	41.72	62.24
MAUS	64.13	75.16	68.36	83.45	57.99
AffectiveROAD	**66.82**	**74.92**	**73.67**	**76.21**	**67.38**
E4 (Baseline)	69.51	77.26	75.00	79.66	70.90

TABLE X: PERFORMANCE OF CNN TRAINED ON DIFFERENT DATASETS AND TESTED ON E4

Training Dataset	ACC (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUC (%)
Combined	66.82	71.38	58.33	73.67	65.59
UTD	58.07	53.10	67.31	62.22	65.19
WESAD	41.93	17.59	87.18	28.25	57.23
MAUS	61.66	77.24	32.69	72.37	61.14
AffectiveROAD	**67.71**	**64.83**	**73.08**	**72.31**	**72.42**
E4 (Baseline)	67.04	84.48	34.62	76.92	70.16

V. Future Outlook and Discussion

Although significant progress has been made in EDA-based stress detection research, our empirical analysis reveals that the field still faces multiple challenges on its path to real-world application. Among these, model accuracy and generalization ability are the most critical bottlenecks at present. The experimental results show that there is considerable room for improvement in the performance of both multi-domain fusion models and cross-domain transfer models. There may be several reasons behind this. First, at the **data processing pipeline** level, although advanced decomposition algorithms like cvxEDA already possess a certain degree of robustness, existing research generally lacks a systematic exploration of preprocessing strategies. More importantly, using a fixed-length analysis window (e.g., 30 seconds) limits the model's flexibility, as effective prediction may only require a few seconds of data, while at other times, a longer time series might be needed to capture complex physiological dynamics. To address this issue, some studies have begun to explore dynamic window adjustment strategies, such as training an agent through deep reinforcement learning to trade off between prediction accuracy and response latency, thereby dynamically optimizing the window length—a highly valuable idea for future reference [183].

Second, **significant physiological differences between individuals** are a key factor limiting the generalization ability of models. Current unimodal EDA research has paid insufficient attention to this issue, with some work only going as far as roughly grouping users through unsupervised clustering. To build truly effective personalized models, we should draw from the more mature experiences in the multimodal domain and actively explore strategies such as transfer learning, model fine-tuning, and using individual static physiological data (e.g., baseline heart rate, age) as model inputs to achieve precise adaptation to individual response patterns [184]. Closely related to individual differences is the domain shift problem caused by **scenario diversity**. Our cross-domain experiments clearly confirm that a model trained in a specific scenario (e.g., a laboratory) is difficult to apply directly to another scenario (e.g., real-world driving). Although data fusion is an intuitive solution, our experiments show that simple aggregation may even lead to worse performance than a model trained on a single, high-ecological-validity dataset (like AffectiveROAD) by ”diluting” the effective signals of a specific domain. Therefore, the focus of future research should shift to more advanced model-level solutions, such as using methods like Maximum Mean Discrepancy (MMD) to align the feature distributions of different domains to learn domain-invariant feature representations [185], or using contrastive learning and related alignment algorithms (CORAL) to bring different domains closer in the feature space [186]. The research by [187] is a valuable attempt, as it combines supervised contrastive learning with metadata, enabling the model to learn a general stress representation that is independent of specific users while also being able to adapt to individual differences.

Data-level challenges are equally severe. On one hand, there is the problem of **data scarcity**. Although this paper has systematically compiled 21 public datasets, their scale and diversity are still insufficient for training complex deep learning models capable of covering a wide range of scenarios. Data generation techniques offer an effective solution. In addition to the already preliminarily validated method of synthetic data based on physiological signal models [132], the application potential of more advanced generative models, such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models [188], in this field is yet to be fully explored. [189] has successfully designed a conditional GAN that prevents mode collapse by introducing diversity constraints, achieving the generation of high-quality EDA and skin temperature (ST) signals, providing a paradigm for future research. On the other hand, there is the issue of **label quality**. Current research generally relies on activity labels from experimental protocols to define stress states, but this overlooks confounding factors such as individual differences and physical activity, leading to a decline in model performance in the real world [190]. While using salivary cortisol as a gold standard can provide more objective labels, its invasiveness and high cost limit its application in mobile scenarios [191]; relying on user self-reports is prone to subjective bias and may lead to "reporting fatigue."

To break free from the dependence on large-scale, high-quality labeled data, Self-Supervised Learning (SSL) shows enormous potential. SSL learns universal feature representations from massive amounts of unlabeled data by designing ingenious pretext tasks. [192] proposed a general framework that successfully learns high-level features from unlabeled sensor data through various pretext tasks, such as signal transformation recognition. Although this framework is highly general, it lacks specific optimization for particular signals. Subsequent research has begun to focus on the characteristics of physiological signals, for example, by learning the dynamic changes of physiological baselines through time-series prediction tasks [193], [194]. Among these, the work by [195] is particularly outstanding. By tailoring data augmentation strategies for the EDA signal (e.g., altering frequency components, distorting tonic and phasic components) and combining them with the SimCLR contrastive learning framework, the learned representations comprehensively outperformed traditional supervised learning methods on downstream tasks, opening up a new research paradigm for the field.

As models move towards practical deployment, a series of **engineering and ethical challenges** must also be addressed. First is **data privacy and security**. As highly sensitive personal data, physiological signals pose a huge risk of leakage when stored centrally. Federated Learning (FL) provides an effective solution by allowing models to be trained on users' local devices without the data ever leaving, aggregating global knowledge only by exchanging encrypted model parameters. This achieves model optimization while protecting user privacy [196], [197]. Of course, the potential performance degradation of models due to FL must also be noted [198]. Second is **model deployment efficiency**. The computational and storage resources of wearable and other edge devices are extremely limited, requiring models to be lightweight. [133] utilized Neural Architecture Search (NAS) technology to automatically design the lightweight TinyStressNet model, which significantly reduces computational overhead while maintaining performance comparable to that of large models. In addition, biologically inspired Spiking Neural Networks (SNNs) also show great potential due to their low power consumption characteristics, with related research having confirmed that they can save tens of times more energy compared to traditional Artificial Neural Networks (ANNs) [199]. Finally, there is the issue of **device heterogeneity**. Current datasets mostly rely on professional research-grade devices (like the Empatica E4). Whether models can maintain consistent performance across consumer-grade wearable devices from different brands is key to determining whether the technology can be widely adopted. Some studies have already begun to investigate this reproducibility issue, laying the groundwork for future research on cross-device model generalization [200].

Finally, as model complexity increases, **interpretability** becomes crucial. The "black-box" nature of deep learning models hinders their application in high-risk fields like healthcare, as we need to understand the physiological basis on which a model makes its judgments. eXplainable AI (XAI) provides powerful tools for this purpose. [201] used the Integrated Gradients method to quantify the contribution of each point in a time series to the prediction result. Their analysis intuitively showed that the rising and recovery phases of the EDA signal's peaks contribute most to stress prediction, which is in complete agreement with existing physiological knowledge. Similarly, [202] used SHAP values for post-hoc analysis of the feature importance of traditional machine learning models, verifying the consistency of model decisions with statistical analysis and enhancing the model's transparency. Taking a step further, [203] introduced Physics-Informed Neural Networks (PINNs) to the field. By embedding the dynamic differential equations of the EDA signal as a constraint in the loss function, the model learns in a data-driven manner while adhering to physiological laws. The internal parameters it learns (such as decay rates) thus acquire clear physiological meaning, providing a new direction for building truly trustworthy stress detection models.

VI. Conclusion

This paper presents the first end-to-end systematic survey and rigorous cross-scenario empirical benchmark for unimodal stress detection using wearable EDA sensors. We systematically reviewed the entire technical pipeline from data processing to modeling and consolidated 21 public datasets to address the critical issues of methodological fragmentation and resource scarcity. The core contribution is a novel empirical benchmark established by evaluating mainstream ML and DL models across five diverse datasets, which quantifies the core challenges of multi-domain data fusion and cross-domain generalization. Our experiments provide a key empirical insight: end-to-end DL models demonstrate significantly superior robustness and generalization capabilities over traditional ML approaches that rely on handcrafted features. Critically, we reveal that the generalization performance is strongly correlated with the ecological validity of the training data; a model trained on a real-world driving dataset, for instance, can successfully transfer its knowledge to an unseen target scenario and even outperform the baseline model. This work provides an essential performance benchmark and empirically-grounded guidance for model selection, aiming to standardize methodologies, enhance reproducibility, and accelerate the translation of EDA-based stress detection from laboratory to real-world applications.

While this study provides a foundational benchmark, we acknowledge its limitations, which in turn illuminate promising directions for future research. Our evaluation was primarily focused on mainstream supervised learning models and did not extend to more advanced paradigms such as SSL or sophisticated domain adaptation techniques; furthermore, a deep exploration of personalized models to account for individual physiological differences was beyond the scope of this work. Future research should therefore prioritize the investigation of SSL and domain adaptation algorithms to enhance model generalization across diverse scenarios, devices, and individuals. We also advocate for research into dynamic windowing strategies, advanced personalized modeling techniques such as transfer learning and fine-tuning, and the application of XAI. Integrating these approaches will be crucial for developing transparent, trustworthy, and effective stress detection models that are ready for reliable, real-world deployment in health applications.

References

[1] J. B. Kirkbride, D. M. Anglin, I. Colman, J. Dykxhoorn, P. B. Jones, P. Patalay, A. Pitman, E. Soneson, T. Steare, T. Wright, and S. L. Griffiths, "The social determinants of mental health and disorder: evidence, prevention and recommendations," *World Psychiatry*, vol. 23, no. 1, pp. 58–90, Feb. 2024.

[2] H. Selye, "The general adaptation syndrome and the diseases of adaptation," *J. Clin. Endocrinol. Metab.*, vol. 6, no. 2, pp. 117–230, Feb. 1946.

[3] A. Mariotti, "The effects of chronic stress on health: new insights into the molecular mechanisms of brain-body communication," *Future Sci. OA*, vol. 1, no. 3, Nov. 2015.

[4] World Health Organization, "Over a billion people living with mental health conditions," Sep 2025.

[5] W. Van Zoonen and A. E. Sivunen, "The impact of remote work and mediated communication frequency on isolation and psychological distress," *Eur. J. Work Organ. Psychol.*, vol. 31, no. 4, pp. 610–621, 2022.

[6] M. L. G. Ramírez, J. P. G. Vázquez, M. D. Rodríguez, L. A. Padilla-López, G. M. Galindo-Aldana, and D. Cuevas-González, "Wearables for Stress Management: A Scoping Review," *Healthcare*, vol. 11, no. 17, Aug. 2023.

[7] K. M. Harris, A. E. Gaffey, J. E. Schwartz, D. S. Krantz, and M. M. Burg, "The Perceived Stress Scale as a Measure of Stress: Decomposing Score Variance in Longitudinal Behavioral Medicine Studies," *Ann. Behav. Med.*, vol. 57, no. 10, pp. 846–854, Sep. 2023.

[8] T. Iqbal, A. Elahi, W. Wijns, and A. Shahzad, "Cortisol detection methods for stress monitoring in connected health," *Health Sci. Rev.*, vol. 6, 2023.

[9] S. Gedam and S. Paul, "A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques," *IEEE Access*, vol. 9, pp. 84 045–84 066, 2021.

[10] H. G. Kim, E. J. Cheon, D. S. Bai, Y. H. Lee, and B. H. Koo, "Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature," *Psychiatry Investig.*, vol. 15, no. 3, pp. 235–245, Mar. 2018.

[11] J. Yedukondalu, K. Sunkara, V. Radhika, S. Kondaveeti, M. Anumothu, and M. K. Yadadavalli, "Cognitive load detection through EEG lead wise feature optimization and ensemble classification," *Sci. Rep.*, vol. 15, no. 1, Jan. 2025.

[12] M. Namvari, J. Lipoth, S. Knight, A. A. Jamali, M. Hedayati, R. J. Spiteri, and S. Syed-Abdul, "Photoplethysmography Enabled Wearable Devices and Stress Detection: A Scoping Review," *J. Pers. Med.*, vol. 12, no. 11, 2022.

[13] S. A. Patil and A. N. Paithane, "AI-Driven Multimodal Stress Detection: A Comparative Study," *Biomedical and Pharmacology Journal*, vol. 18, no. Special Issue, Mar. 2025.

[14] L. Zhu, P. Spachos, P. C. Ng, Y. Yu, Y. Wang, K. Plataniotis, and D. Hatzinakos, "Stress detection through wrist-based electrodermal activity monitoring and machine learning," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 5, pp. 2155–2165, May 2023.

[15] W. Boucsein, *Electrodermal Activity*. Boston, MA: Springer US, 2012.

[16] L. Zhu, P. Spachos, and S. Gregori, "Multimodal physiological signals and machine learning for stress detection by wearable devices," in *Proc. 2022 IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2022, pp. 1–6.

[17] A. Singh, A. Kumar, K. Singh, and S. Kumar, "A deep learning approach for early stress detection using electrodermal activity through wearable devices," in *Pattern Recognition. ICPR 2024 Int. Workshops Challenges*, S. Palaiahnakote, S. Schuckers, J.-M. Ogier, P. Bhattacharya, U. Pal, and S. Bhattacharya, Eds. Cham: Springer Nature Switzerland, 2025, pp. 49–66.

[18] R. Holder, R. K. Sah, M. Cleveland, and H. Ghasemzadeh, "Comparing the predictability of sensor modalities to detect stress from wearable sensor data," in *Proc. 2022 IEEE 19th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2022, pp. 557–562.

[19] K. Wang and P. Guo, "An ensemble classification model with unsupervised representation learning for driving stress recognition using physiological signals," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3303–3315, Jun. 2021.

[20] Y. Guo, K. Yang, and Y. Wu, "A multi-modality attention network for driver fatigue detection based on frontal EEG, EDA and PPG signals," *IEEE J. Biomed. Health Inform.*, vol. 29, no. 6, pp. 4009–4022, Jun. 2025.

[21] B. Choi, H. Jebelli, and S. Lee, "Feasibility analysis of electrodermal activity (EDA) acquired from wearable sensors to assess construction workers' perceived risk," *Saf. Sci.*, vol. 115, pp. 110–120, 2019.

[22] M.-B. Hossain, K. Golzari, Y. Kong, B. J. Derrick, R. E. Moon, M. J. Natoli, M. C. Ellis, C. Winstead-Derlega, S. I. Gonzalez, C. M. Allen, M. S. Makowski, B. M. Keuski, J. J. Freiberger, H. F. Posada-Quintero, and K. H. Chon, "Prediction of central nervous system

oxygen toxicity symptoms using electrodermal activity and machine learning," *Biocybern. Biomed. Eng.*, vol. 44, no. 2, pp. 304–311, 2024.

[23] V. Abromavičius, A. Serackis, A. Katkevičius, M. Kazlauskas, and T. Sledevič, "Prediction of exam scores using a multi-sensor approach for wearable exam stress dataset with uniform preprocessing," *Technol. Health Care*, vol. 31, no. 6, pp. 2499–2511, 2023.

[24] H. Lee and A. Kleinsmith, "Public speaking anxiety in a real classroom: Towards developing a reflection system," in *Extended Abstr. 2019 CHI Conf. Human Factors Comput. Syst. (CHI EA '19)*. New York, NY, USA: ACM, 2019, pp. 1–6.

[25] J. Gorson, K. Cunningham, M. Worsley, and E. O'Rourke, "Using electrodermal activity measurements to understand student emotions while programming," in *Proc. 2022 ACM Conf. Int. Comput. Educ. Res. - vol. 1 (ICER '22)*. New York, NY, USA: ACM, 2022, pp. 105–119.

[26] V. Abromavičius, A. Serackis, A. Katkevičius, M. Kazlauskas, and T. Sledevič, "Prediction of exam scores using a multi-sensor approach for wearable exam stress dataset with uniform preprocessing," *Technol. Health Care*, vol. 31, no. 6, pp. 2499–2511, 2023.

[27] A. Hsu, "Quantifying exam stress progressions using electrodermal activity and machine learning," in *Proc. 2023 IEEE 23rd Int. Conf. Bioinform. Bioeng. (BIBE)*, 2023, pp. 434–438.

[28] G. M. Machado and A. Soni, "Predicting academic performance: A comprehensive electrodermal activity study," in *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*, A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, and I. I. Bittencourt, Eds. Cham: Springer Nature Switzerland, 2024, pp. 140–148.

[29] P. Sariñana-González, Á. Romero-Martínez, and L. Moya-Albiol, "Cooperation induces an increase in emotional response, as measured by electrodermal activity and mood," *Curr. Psychol.*, vol. 36, no. 2, pp. 366–375, Jun. 2017.

[30] P. E. Fortin, E. Sulmont, and J. Cooperstock, "Detecting perception of smartphone notifications using skin conductance responses," in *Proc. 2019 CHI Conf. Human Factors Comput. Syst. (CHI '19)*. New York, NY, USA: ACM, 2019, pp. 1–9.

[31] E. Babaei, B. Tag, T. Dingler, and E. Velloso, "A critique of electrodermal activity practices at CHI," in *Proc. 2021 CHI Conf. Human Factors Comput. Syst. (CHI '21)*. New York, NY, USA: ACM, 2021, pp. 177:1–177:14.

[32] J. Sehrt, U. Yilmaz, T. Kosch, and V. Schwind, "Closing the loop: The effects of biofeedback awareness on physiological stress response using electrodermal activity in virtual reality," in *Extended Abstr. CHI Conf. Human Factors Comput. Syst. (CHI EA '24)*. New York, NY, USA: ACM, 2024, pp. 76:1–76:7.

[33] O. N. Rahma, A. P. Putra, A. Rahmatillah, Y. S. K. A. Putri, N. D. Fajriaty, K. Ain, and R. Chai, "Electrodermal activity for measuring cognitive and emotional stress level," *J. Med. Signals Sens.*, vol. 12, no. 2, pp. 155–162, May 2022.

[34] J. Shukla, M. Barreda-Ángeles, J. Oliver, G. C. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," *IEEE Trans. Affect. Comput.*, vol. 12, no. 4, pp. 857–869, Oct. 2021.

[35] H. F. Posada-Quintero and K. H. Chon, "Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review," *Sensors*, vol. 20, no. 2, 2020.

[36] R. Sánchez-Reolid, M. T. López, and A. Fernández-Caballero, "Machine Learning for Stress Detection from Electrodermal Activity: A Scoping Review," *Preprints*, Nov. 2020.

[37] A. Horvers, N. Tombeng, T. Bosse, A. W. Lazonder, and I. Molenaar, "Detecting Emotions through Electrodermal Activity in Learning Contexts: A Systematic Review," *Sensors*, vol. 21, no. 23, 2021.

[38] G. Vos, K. Trinh, Z. Sarnyai, and M. Rahimi Azghadi, "Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review," *Int. J. Med. Informat.*, vol. 173, p. 105026, 2023.

[39] A. Klimek, I. Mannheim, G. Schouten, E. J. M. Wouters, and M. W. H. Peeters, "Wearables measuring electrodermal activity to assess perceived stress in care: a scoping review," *Acta Neuropsychiatr.*, vol. 37, Mar. 2023.

[40] M. Kyrou, I. Kompatsiaris, and P. C. Petrantonakis, "Deep Learning Approaches for Stress Detection: A Survey," *IEEE Trans. Affect. Comput.*, vol. 16, no. 2, pp. 499–517, Apr. 2025.

[41] J. Kallio, E. Vildjiounaite, J. Tervonen, and M. B. López, "A survey on sensor-based techniques for continuous stress monitoring in knowledge work environments," *ACM Trans. Comput. Healthcare*, vol. 6, no. 3, pp. 36:1–36:31, May 2025.

[42] M. Gjoreski, T. Kolenik, T. Knez, M. Luštrek, M. Gams, H. Gjoreski, and V. Pejović, "Datasets for cognitive load inference using wearable sensors and psychological traits," *Appl. Sci.*, vol. 10, no. 11, p. 3843, 2020.

[43] A. Ometov, A. Mezina, H.-C. Lin, O. Arponen, R. Burget, and J. Nurmi, "Stress and emotion open access data: A review on datasets, modalities, methods, challenges, and future research perspectives," *J. Healthc. Inform. Res.*, vol. 9, no. 3, pp. 247–279, Sep. 2025.

[44] J. P. C. J. Pegueros and J. Rodríguez-Arce, "Physiological datasets in stress and anxiety research: A systematic review," *Biomed. Signal Process. Control*, vol. 108, p. 107928, 2025.

[45] M. Migovich, D. Adiani, M. Breen, A. Swanson, T. J. Vogus, and N. Sarkar, "Stress detection of autistic adults during simulated job interviews using a novel physiological dataset and machine learning," *ACM Trans. Access. Comput.*, vol. 17, no. 1, pp. 2:1–2:25, Mar. 2024.

[46] M. Laporte, E. Fu, M. Gjoreski, and M. Langheinrich, "Introduction of the LAUREATE dataset: The longitudinal multimodal student experience dataset for affect and memory research," in *Adjunct Proc. 2022 ACM Int. Joint Conf. Pervasive Ubiquitous Comput. 2022 ACM Int. Symp. Wearable Comput. (UbiComp/ISWC '22 Adjunct)*. New York, NY, USA: ACM, 2023, pp. 67–69.

[47] M. Benchekroun, D. Istrate, V. Zalc, and D. Lenne, "MMSD: A multi-modal dataset for real-time, continuous stress detection from physiological signals," in *Proc. 15th Int. Joint Conf. Biomed. Eng. Syst. Technol. (BIOSTEC 2022) - vol. 5: HEALTHINF*, Feb. 2022, pp. 240–248.

[48] M. Jaiswal, C.-P. Bara, Y. Luo, M. Burzo, R. Mihalcea, and E. M. Provost, "MuSE: A multimodal dataset of stressed emotion," in *Proc. 12th Lang. Resour. Eval. Conf.* Marseille, France: European Language Resources Association, May 2020, pp. 1499–1510.

[49] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[50] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij, "The swell knowledge work dataset for stress and user modeling research," in *Proc. 16th Int. Conf. Multimodal Interaction (ICMI '14)*. New York, NY, USA: ACM, 2014, pp. 291–298.

[51] J. Birjandtalab, D. Cogan, M. B. Pouyan, and M. Nourani, "A non-eeg biosignals dataset for assessment and visualization of neurological status," in *Proc. 2016 IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2016, pp. 110–114.

[52] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. V. Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interaction (ICMI '18)*. New York, NY, USA: ACM, 2018, pp. 400–408.

[53] N. E. Haouij, J.-M. Poggi, S. Sevestre-Ghalila, R. Ghozi, and M. Jaïdane, "Affectiveroad system and database to assess driver's attention," in *Proc. 33rd Annu. ACM Symp. Appl. Comput. (SAC '18)*. New York, NY, USA: ACM, 2018, pp. 800–803.

[54] V. Markova, T. Ganchev, and K. Kalinkov, "CLAS: A database for cognitive load, affect and stress recognition," in *Proc. Int. Conf. Biomed. Innov. Appl. (BIA-2019)*, 2019.

[55] H. Svoren, V. Thambawita, P. Halvorsen, P. Jakobsen, E. Garcia-Ceja, F. M. Noori, H. L. Hammer, M. Lux, M. A. Riegler, and S. A. Hicks, "Toadstool: A dataset for training emotional intelligent machines playing super mario bros," in *Proc. 11th ACM Multimedia Syst. Conf. (MMSys '20)*. New York, NY, USA: ACM, 2020, pp. 309–314.

[56] W.-K. Beh, Y.-H. Wu, and A.-Y. Wu, "MAUS: A dataset for mental workload assessment on n-back task using wearable sensor," arXiv preprint arXiv:2111.02561, 2021. [Online]. Available: https://arxiv.org/abs/2111.02561

[57] M. Yadav, M. N. Sakib, E. H. Nirjhar, K. Feng, A. H. Behzadan, and T. Chaspari, “Exploring individual differences of public speaking anxiety in real-life and virtual presentations,” *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1168–1182, Jul. 2022.

[58] M. Hosseini, F. Sohrab, R. Gottumukkala, R. T. Bhupatiraju, S. Katragadda, J. Raitoharju, A. Iosifidis, and M. Gabbouj, “EmpathicSchool: A multimodal dataset for real-time facial expressions and physiological data analysis under different stress conditions,” arXiv preprint arXiv:2209.13542, 2022. [Online]. Available: https://arxiv.org/abs/2209.13542

[59] M. R. Amin, D. S. Wickramasuriya, and R. T. Faghih, “A wearable exam stress dataset for predicting grades using physiological signals,” in *Proc. 2022 IEEE Healthc. Innov. Point Care Technol. (HI-POCT)*, Mar. 2022, pp. 30–36.

[60] S. Hosseini, R. Gottumukkala, S. Katragadda, R. T. Bhupatiraju, Z. Ashkar, C. W. Borst, and K. Cochran, “A multimodal sensor dataset for continuous stress detection of nurses in a hospital,” *Sci. Data*, vol. 9, no. 1, p. 255, Jun. 2022.

[61] H. Chaptoukaev, V. Strizhkova, M. Panariello, B. Dalpaos, A. Reka, V. Manera, S. Thummler, E. Ismailova, N. W., F. Bremond, M. Todisco, M. A. Zuluaga, and L. M. Ferrari, “StressID: A multimodal dataset for stress identification,” in *Proc. 37th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track*, 2023. [Online]. Available: https://openreview.net/forum?id=qWsQi9DGJb

[62] R. M. Sabour, Y. Benezeth, P. D. Oliveira, J. Chappé, and F. Yang, “UBFC-Phys: A multimodal database for psychophysiological studies of social stress,” *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 622–636, Jan. 2023.

[63] Ç. Çöpürkaya, E. Meriç, E. B. Erik, B. Kocaçınar, F. P. Akbulut, and C. Catal, “Investigating the effects of stress on achievement: BIOSTRESS dataset,” *Data Brief*, vol. 49, p. 109297, 2023.

[64] S. Kang, W. Choi, C. Y. Park, N. Cha, A. Kim, A. H. Khandoker, L. Hadjileontiadis, H. Kim, Y. Jeong, and U. Lee, “K-EmoPhone: A mobile and wearable dataset with in-situ emotion, stress, and attention labels,” *Sci. Data*, vol. 10, no. 1, p. 351, Jun. 2023.

[65] A. Heimerl, P. Prajod, S. Mertes, T. Baur, M. Kraus, A. Liu, H. Risack, N. Rohleder, E. André, and L. Becker, “ForDigitStress: A multi-modal stress dataset employing a digital job interview scenario,” arXiv preprint arXiv:2303.07742, 2023. [Online]. Available: https://arxiv.org/abs/2303.07742

[66] A. Hashky, B. Rheault, A. R. Ahsan, B. Benda, T. Audino, S. Lonneman, and E. D. Ragan, “Multi-modal user modeling for task guidance: A dataset for real-time assistance with stress and interruption dynamics,” in *Proc. 2024 IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops (VRW)*, Mar. 2024, pp. 544–550.

[67] S. Campanella, A. Altaleb, A. Belli, P. Pierleoni, and L. Palma, “PPG and EDA dataset collected with empatica e4 for stress assessment,” *Data Brief*, vol. 53, p. 110102, 2024.

[68] G. Dogan, F. P. Akbulut, and C. Catal, “Biosignals, facial expressions, and speech as measures of workplace stress: Workstress3d dataset,” *Data Brief*, vol. 54, p. 110303, Mar. 2024.

[69] S. Das, N. L. Lund, C. R. González, and L. H. Clemmensen, “EmoPairCompete - physiological signals dataset for emotion and frustration assessment under team and competitive behaviors,” in *Proc. ICLR 2024 Workshop Learn. Time Series Health*, 2024. [Online]. Available: https://openreview.net/forum?id=BvgAzJX40Z

[70] C. Setz, B. Arnrich, J. Schumm, R. L. Marca, G. Tröster, and U. Ehlert, “Discriminating stress from cognitive load using a wearable EDA device,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 410–417, Mar. 2010.

[71] J. Bakker, M. Pechenizkiy, and N. Sidorova, “What’s your current stress level? detection of stress patterns from GSR sensor data,” in *Proc. 2011 IEEE 11th Int. Conf. Data Mining Workshops*, Dec. 2011, pp. 573–580.

[72] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, “Ambulatory stress monitoring with minimally-invasive wearable sensors,” Dept. Comput. Sci. Eng., Texas A&M Univ., Tech. Rep. tamu-cs-tr-2010-11-1, 2010.

[73] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, “Stress detection from speech and galvanic skin response signals,” in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 209–214.

[74] J. Hernandez, R. R. Morris, and R. W. Picard, “Call center stress recognition with person-specific models,” in *Affective Computing and Intelligent Interaction*, S. D’Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 125–134.

[75] M. V. Villarejo, B. G. Zapirain, and A. M. Zorrilla, “A stress sensor based on galvanic skin response (GSR) controlled by ZigBee,” *Sensors*, vol. 12, no. 5, pp. 6075–6101, May 2012.

[76] A. Greco, A. Lanata, L. Citi, N. Vanello, G. Valenza, and E. P. Scilingo, “Skin admittance measurement for emotion recognition: A study over frequency sweep,” *Electronics*, vol. 5, no. 3, p. 46, 2016.

[77] R. Zangróniz, A. Martínez-Rodrigo, J. M. Pastor, M. T. López, and A. Fernández-Caballero, “Electrodermal activity sensor for classification of calm/distress condition,” *Sensors*, vol. 17, no. 10, p. 2324, 2017.

[78] J. Hernandez, R. R. Morris, and R. W. Picard, “Call center stress recognition with person-specific models,” in *Affective Computing and Intelligent Interaction*, S. D’Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 125–134.

[79] J. Hernandez, I. Riobo, A. Rozga, G. D. Abowd, and R. W. Picard, “Using electrodermal activity to recognize ease of engagement in children during social interactions,” in *Proc. 2014 ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp ’14)*. New York, NY, USA: ACM, 2014, pp. 307–317.

[80] W. Chen, N. Jaques, S. Taylor, A. Sano, S. Fedor, and R. W. Picard, “Wavelet-based motion artifact removal for electrodermal activity,” in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2015, pp. 6223–6226.

[81] C. Tronstad, O. M. Staal, S. Sælid, and Ø. G. Martinsen, “Model-based filtering for artifact and noise suppression with state estimation for electrodermal activity measurements in real time,” in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2750–2753.

[82] J. Shukla, M. Barreda-Ángeles, J. Oliver, and D. Puig, “Efficient wavelet-based artifact removal for electrodermal activity in real-world applications,” *Biomed. Signal Process. Control*, vol. 42, pp. 45–52, 2018.

[83] P. Zontone, A. Affanni, R. Bernardini, A. Piras, and R. Rinaldo, “Stress detection through electrodermal activity (EDA) and electrocardiogram (ECG) analysis in car drivers,” in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2019, pp. 1–5.

[84] P. Zontone, A. Affanni, R. Bernardini, L. D. Linz, A. Piras, and R. Rinaldo, “Analysis of physiological signals for stress recognition with different car handling setups,” *Electronics*, vol. 11, no. 6, p. 888, 2022.

[85] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, “Automatic identification of artifacts in electrodermal activity data,” in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2015, pp. 1934–1937.

[86] Y. Zhang, M. Haghdan, and K. S. Xu, “Unsupervised motion artifact detection in wrist-measured electrodermal activity data,” in *Proc. 2017 ACM Int. Symp. Wearable Comput. (ISWC ’17)*. New York, NY, USA: ACM, 2017, pp. 54–57.

[87] I. R. Kleckner, R. M. Jones, O. Wilder-Smith, J. B. Wormwood, M. Akcakaya, K. S. Quigley, C. Lord, and M. S. Goodwin, “Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 7, pp. 1460–1467, Jul. 2018.

[88] D. R. Bach, G. Flandin, K. J. Friston, and R. J. Dolan, “Time-series analysis for rapid event-related skin conductance responses,” *J. Neurosci. Methods*, vol. 184, no. 2, pp. 224–234, Nov. 2009.

[89] ——, “Modelling event-related skin conductance responses,” *Int. J. Psychophysiol.*, vol. 75, no. 3, pp. 349–356, 2010.

[90] D. R. Bach, J. Daunizeau, N. Kuelzow, K. J. Friston, and R. J. Dolan, “Dynamic causal modeling of spontaneous fluctuations in skin conductance,” *Psychophysiology*, vol. 48, no. 2, pp. 252–257, 2011.

[91] M. Benedek and C. Kaernbach, “A continuous measure of phasic electrodermal activity,” *J. Neurosci. Methods*, vol. 190, no. 1, pp. 80–91, 2010.

[92] ——, "Decomposition of skin conductance data by means of nonnegative deconvolution," *Psychophysiology*, vol. 47, no. 4, pp. 647–658, 2010.

[93] D. R. Bach, K. J. Friston, and R. J. Dolan, "Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations," *Int. J. Psychophysiol.*, vol. 76, no. 1, pp. 52–55, Apr. 2010.

[94] D. R. Bach, J. Daunizeau, N. Kuelzow, K. J. Friston, and R. J. Dolan, "Dynamic causal modeling of spontaneous fluctuations in skin conductance," *Psychophysiology*, vol. 48, no. 2, pp. 252–257, Feb. 2011.

[95] D. R. Bach, K. J. Friston, and R. J. Dolan, "An improved algorithm for model-based analysis of evoked skin conductance responses," *Biol. Psychol.*, vol. 94, no. 3, pp. 490–497, Dec. 2013.

[96] T. Chaspari, A. Tsiartas, L. I. Stein, S. A. Cermak, and S. S. Narayanan, "Sparse representation of electrodermal activity with knowledge-driven dictionaries," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 3, pp. 960–971, Mar. 2015.

[97] D. R. Bach and M. Staib, "A matching pursuit algorithm for inferring tonic sympathetic arousal from spontaneous skin conductance fluctuations," *Psychophysiology*, vol. 52, no. 8, pp. 1106–1112, Aug. 2015.

[98] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, Apr. 2016.

[99] H. F. Posada-Quintero, J. P. Florian, Á. D. Orjuela-Cañón, and K. H. Chon, "Highly sensitive index of sympathetic activity based on time-frequency spectral analysis of electrodermal activity," *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, vol. 311, no. 3, pp. R582–R591, 2016.

[100] S. Jain, U. Oswal, K. S. Xu, B. Eriksson, and J. Haupt, "A compressed sensing based decomposition of electrodermal activity signals," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2142–2151, Sep. 2017.

[101] F. Hernando-Gallego, D. Luengo, and A. Artés-Rodríguez, "Feature extraction of galvanic skin responses by nonnegative sparse deconvolution," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1385–1394, Sep. 2018.

[102] M. R. Amin and R. T. Faghih, "Sparse deconvolution of electrodermal activity via continuous-time system identification," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 9, pp. 2585–2595, Sep. 2019.

[103] ——, "Tonic and phasic decomposition of skin conductance data: A generalized-cross-validation-based block coordinate descent approach," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 745–749.

[104] R. Amin and R. T. Faghih, "Physiological characterization of electrodermal activity enables scalable near real-time autonomic nervous system activation inference," *PLoS Comput. Biol.*, vol. 18, no. 7, p. e1010275, Jul. 2022.

[105] S. Alam, M. R. Amin, and R. T. Faghih, "Sparse multichannel decomposition of electrodermal activity with physiological priors," *IEEE Open J. Eng. Med. Biol.*, vol. 4, pp. 234–250, 2023.

[106] Ledalab Team, "Ledalab - a toolbox for skin conductance data analysis," Online, 2025. [Online]. Available: http://www.ledalab.de/

[107] SourceForge, "SCRalyze - skin conductance response analysis tool," Online, 2025. [Online]. Available: https://sourceforge.net/projects/scralyze/

[108] PsPM Team, "PsPM - the psychophysiological modeling toolbox," Online, 2025. [Online]. Available: https://sourceforge.net/projects/pspm/

[109] F. H. Gallego, "sparsEDA: A sparse decomposition approach for electrodermal activity," Online, 2025. [Online]. Available: https://github.com/fhernandogallego/sparsEDA

[110] A. Greco, G. Valenza, and E. P. Scilingo, "cvxEDA: A convex optimization approach to electrodermal activity processing," Online, 2025. [Online]. Available: https://github.com/lciti/cvxEDA

[111] D. R. Bach, "A head-to-head comparison of SCRalyze and Ledalab, two model-based methods for skin conductance analysis," *Biol. Psychol.*, vol. 103, pp. 63–68, 2014.

[112] E. Lutin, R. Hashimoto, W. D. Raedt, and C. V. Hoof, "Feature extraction for stress detection in electrodermal activity," in *Proc. 14th Int. Joint Conf. Biomed. Eng. Syst. Technol. - vol. 4: BIOSIGNALS*. SCITEPRESS, 2021, pp. 177–185.

[113] Y. R. Veeranki, N. Ganapathy, R. Swaminathan, and H. F. Posada-Quintero, "Comparison of electrodermal activity signal decomposition techniques for emotion recognition," *IEEE Access*, vol. 12, pp. 19 952–19 966, 2024.

[114] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behav. Res. Methods*, vol. 53, no. 4, pp. 1689–1696, Feb. 2021.

[115] P. Bota, R. Silva, C. Carreiras, A. Fred, and H. P. da Silva, "BioSPPy: A Python toolbox for physiological signal processing," *SoftwareX*, vol. 26, p. 101712, 2024.

[116] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon, "Power spectral density analysis of electrodermal activity for sympathetic function assessment," *Ann. Biomed. Eng.*, vol. 44, no. 10, pp. 3124–3135, Oct. 2016.

[117] R. Martinez, A. Salazar-Ramirez, A. Arruti, E. Irigoyen, J. I. Martin, and J. Muguerza, "A self-paced relaxation response detection system based on galvanic skin response analysis," *IEEE Access*, vol. 7, pp. 43 730–43 741, 2019.

[118] M. Nardelli, A. Greco, L. Sebastiani, and E. P. Scilingo, "ComEDA: A new tool for stress assessment based on electrodermal activity," *Comput. Biol. Med.*, vol. 150, p. 106144, 2022.

[119] L. Lavezzo, A. Gargano, E. P. Scilingo, and M. Nardelli, "Zooming into the complex dynamics of electrodermal activity recorded during emotional stimuli: A multiscale approach," *Bioengineering*, vol. 11, no. 6, p. 520, 2024.

[120] P. Ghaderyan and A. Abbasi, "An efficient automatic workload estimation method based on electrodermal activity using pattern classifier combinations," *Int. J. Psychophysiol.*, vol. 110, pp. 91–101, 2016.

[121] A. S. Anusha, J. Joy, S. P. Preejith, J. Joseph, and M. Sivaprakasam, "Differential effects of physical and psychological stressors on electrodermal activity," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 4549–4552.

[122] Y. Liu and S. Du, "Psychological stress level detection based on electrodermal activity," *Behav. Brain Res.*, vol. 341, pp. 50–53, Apr. 2018.

[123] C.-P. Hsieh, Y.-T. Chen, W.-K. Beh, and A.-Y. A. Wu, "Feature selection framework for XGBoost based on electrodermal activity in stress detection," in *Proc. 2019 IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2019, pp. 330–335.

[124] A. A. S., S. P., S. V., S. S. Kumar, A. Shyam, T. J. Akl, S. P. P., and M. Sivaprakasam, "Electrodermal activity based pre-surgery stress detection using a wrist wearable," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 1, pp. 92–100, Jan. 2020.

[125] R. Sánchez-Reolid, A. Martínez-Rodrigo, M. T. López, and A. Fernández-Caballero, "Deep support vector machines for the identification of stress condition from electrodermal activity," *Int. J. Neural Syst.*, vol. 30, no. 7, p. 2050031, Jul. 2020.

[126] C. M. D. Acevedo, J. K. C. Gómez, and C. A. A. Rojas, "Academic stress detection on university students during COVID-19 outbreak by using an electronic nose and the galvanic skin response," *Biomed. Signal Process. Control*, vol. 68, p. 102756, 2021.

[127] L. Zhu, P. C. Ng, Y. Yu, Y. Wang, P. Spachos, D. Hatzinakos, and K. N. Plataniotis, "Feasibility study of stress detection with machine learning through EDA from wearable devices," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 4800–4805.

[128] E. Hosseini, R. Fang, R. Zhang, A. Parenteau, S. Hang, S. Rafatirad, C. Hostinar, M. Orooji, and H. Homayoun, "A low cost EDA-based stress detection using machine learning," in *Proc. 2022 IEEE Int. Conf. Bioinform. Biomed. (BIBM)*, Dec. 2022, pp. 2619–2623.

[129] A. Greco, G. Valenza, J. Lázaro, J. M. Garzón-Rey, J. Aguiló, C. de la Cámara, R. Bailón, and E. P. Scilingo, "Acute stress state classification based on electrodermal activity modeling," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 788–799, Jan. 2023.

[130] Ž. Stržinar, A. Sanchis, A. Ledezma, O. Sipele, B. Pregelj, and I. Škrjanc, "Stress detection using frequency spectrum analysis of

wrist-measured electrodermal activity," *Sensors*, vol. 23, no. 2, p. 963, 2023.

[131] D. Jaiswal, D. Chatterjee, M. B. s., R. K. Ramakrishnan, and A. Pal, "GSR based generic stress prediction system," in *Adjunct Proc. 2023 ACM Int. Joint Conf. Pervasive Ubiquitous Comput. & 2023 ACM Int. Symp. Wearable Comput. (UbiComp/ISWC '23 Adjunct)*. New York, NY, USA: ACM, 2023, pp. 433–438.

[132] F. Vasile, A. Vizziello, N. Brondino, and P. Savazzi, "Stress state classification based on deep neural network and electrodermal activity modeling," *Sensors*, vol. 23, no. 5, p. 2504, Feb. 2023.

[133] D. Jaiswal, S. Mukhopadhyay, and V. Sharma, "TinyStressNet: On-device stress assessment with wearable sensors on edge devices," in *Proc. 2024 IEEE Int. Conf. Pervasive Comput. Commun. Workshops other Affiliated Events (PerCom Workshops)*, Mar. 2024, pp. 166–171.

[134] L. R. Mercado-Diaz, Y. R. Veeranki, F. Marmolejo-Ramos, and H. F. Posada-Quintero, "EDA-Graph: Graph signal processing of electrodermal activity for emotional states detection," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 8, pp. 4599–4612, Aug. 2024.

[135] S. Sriramprakash, V. D. Prasanna, and O. V. R. Murthy, "Stress detection in working people," *Procedia Comput. Sci.*, vol. 115, pp. 359–366, 2017.

[136] W. Hadi, N. El-Khalili, M. AlNashashibi, G. Issa, and A. A. Al-Banna, "Application of data mining algorithms for improving stress prediction of automobile drivers: A case study in jordan," *Comput. Biol. Med.*, vol. 114, p. 103474, 2019.

[137] F. Delmastro, F. D. Martino, and C. Dolciotti, "Cognitive training and stress detection in MCI frail older people through wearable sensors and machine learning," *IEEE Access*, vol. 8, pp. 65 573–65 590, 2020.

[138] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Continuous stress detection using a wrist device: In laboratory and real life," in *Proc. 2016 ACM Int. Joint Conf. Pervasive Ubiquitous Comput.: Adjunct (UbiComp '16)*. New York, NY, USA: ACM, 2016, pp. 1185–1193.

[139] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Inform.*, vol. 73, pp. 159–170, 2017.

[140] D. Cho, J. Ham, J. Oh, J. Park, S. Kim, N.-K. Lee, and B. Lee, "Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine," *Sensors*, vol. 17, no. 10, p. 2435, 2017.

[141] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, "Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study," *Sensors*, vol. 19, no. 8, p. 1849, 2019.

[142] E. E. Kaczor, S. Carreiro, J. Stapp, B. Chapman, and P. Indic, "Objective measurement of physician stress in the emergency department using a wearable sensor," in *Proc. 53rd Hawaii Int. Conf. Syst. Sci.*, Jan. 2020, pp. 3729–3738.

[143] A. Khalaf, M. Nabian, M. Fan, Y. Yin, J. Wormwood, E. Siegel, K. S. Quigley, L. F. Barrett, M. Akcakaya, C.-A. Chou, and S. Ostadabbas, "Analysis of multimodal physiological signals within and between individuals to predict psychological challenge vs. threat," *Expert Syst. Appl.*, vol. 140, p. 112890, 2020.

[144] Y. S. Can, D. Gokay, D. R. Kılıç, D. Ekiz, N. Chalabianloo, and C. Ersoy, "How laboratory experiments can be exploited for monitoring stress in the wild: A bridge between laboratory and daily life," *Sensors*, vol. 20, no. 3, p. 838, 2020.

[145] E. Eren and T. S. Navruz, "Stress detection with deep learning using BVP and EDA signals," in *Proc. Int. Congr. Human-Comput. Interact., Optim. Robot. Appl. (HORA)*, Jun. 2022, pp. 1–7.

[146] K. Masood and M. A. Alghamdi, "Modeling mental stress using a deep learning framework," *IEEE Access*, vol. 7, pp. 68 446–68 454, 2019.

[147] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2020, pp. 51–57.

[148] R. R. Singh, S. Conjeti, and R. Banerjee, "A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 740–754, 2013.

[149] F. Albertetti, A. Simalastar, and A. Rizzotti-Kaddouri, "Stress detection with deep learning approaches using physiological signals," in *IoT Technologies for HealthCare*, R. Goleva, N. R. d. C. Garcia, and I. M. Pires, Eds. Cham: Springer International Publishing, 2021, pp. 95–111.

[150] A. Liapis, E. Faliagka, C. P. Antonopoulos, G. Keramidas, and N. Voros, "Advancing stress detection methodology with deep learning techniques targeting UX evaluation in AAL scenarios: Applying embeddings for categorical variables," *Electronics*, vol. 10, no. 13, p. 1550, 2021.

[151] S. Eskandar and S. Razavi, "Using deep learning for assessment of workers' stress and overload," in *Proc. 37th Int. Symp. Autom. Robot. Constr. (ISARC)*. IAARC, Oct. 2020, pp. 872–877.

[152] R. Li and Z. Liu, "Stress detection using deep neural networks," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 285, Dec. 2020.

[153] R. K and V. R. M. Oruganti, "Deep multimodal fusion for subject-independent stress detection," in *Proc. 2021 11th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2021, pp. 105–109.

[154] ——, "Stress detection using CNN fusion," in *Proc. TENCON 2021 - 2021 IEEE Region 10 Conf. (TENCON)*, Dec. 2021, pp. 492–497.

[155] R. Kuttala, R. Subramanian, and V. R. M. Oruganti, "Multimodal hierarchical CNN feature fusion for stress detection," *IEEE Access*, vol. 11, pp. 6867–6878, 2023.

[156] A. Kumar, K. Sharma, and A. Sharma, "Hierarchical deep neural network for mental stress state detection using IoT based biomarkers," *Pattern Recognit. Lett.*, vol. 145, pp. 81–87, 2021.

[157] R. K and V. R. M. Oruganti, "Cross domain features for subject-independent stress detection," in *Proc. 2022 IEEE Region 10 Symp. (TENSYMP)*, Jul. 2022, pp. 1–6.

[158] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "A deep residual-based model on multi-branch aggregation for stress and emotion recognition through biosignals," in *Proc. 19th Int. Conf. Electr. Eng./Electron., Comput., Telecommun. Inf. Technol. (ECTI-CON)*, May 2022, pp. 1–4.

[159] Y. Li, K. Li, J. Chen, S. Wang, H. Lu, and D. Wen, "Pilot stress detection through physiological signals using a transformer-based deep learning model," *IEEE Sens. J.*, vol. 23, no. 11, pp. 11 774–11 784, Jun. 2023.

[160] S. Masood, R. Khan, A. A. A. El-Latif, and M. Ahmad, "An FCN-LSTM model for neurological status detection from non-invasive multivariate sensor data," *Neural Comput. Appl.*, vol. 36, no. 1, pp. 77–93, Jan. 2024.

[161] R. Kuttala, R. Subramanian, and V. R. M. Oruganti, "Hierarchical autoencoder frequency features for stress detection," *IEEE Access*, vol. 11, pp. 103 232–103 241, 2023.

[162] J. Lee, H. Lee, and M. Shin, "Driving stress detection using multimodal convolutional neural networks with nonlinear representation of short-term physiological signals," *Sensors*, vol. 21, no. 7, p. 2381, 2021.

[163] M. Gil-Martin, R. San-Segundo, A. Mateos, and J. Ferreiros-Lopez, "Human stress detection with wearable sensors using convolutional neural networks," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 1, pp. 60–70, Jan. 2022.

[164] M. Quadrini, A. Capuccio, D. Falcone, S. Daberdaku, A. Blanda, G. Bellanova, and G. Gerard, "Stress detection with encoding physiological signals and convolutional neural network," *Mach. Learn.*, vol. 113, no. 8, pp. 5655–5683, Aug. 2024.

[165] S. Yang, Y. Gao, Y. Zhu, L. Zhang, Q. Xie, X. Lu, F. Wang, and Z. Zhang, "A deep learning approach to stress recognition through multimodal physiological signal image transformation," *Sci. Rep.*, vol. 15, no. 1, p. 22258, Jul. 2025.

[166] E. Oliver and S. Dakshit, "Cross-modality investigation on WESAD stress classification," arXiv preprint arXiv:2502.18733, 2025. [Online]. Available: https://arxiv.org/abs/2502.18733

[167] R. K. Nath and H. Thapliyal, "Machine learning based anxiety detection in older adults using wristband sensors and context feature," arXiv preprint arXiv:2106.03019, 2021. [Online]. Available: https://arxiv.org/abs/2106.03019

[168] S. Aristizabal, K. Byun, N. Wood, A. F. Mullan, P. M. Porter, C. Campanella, A. Jamrozik, I. Z. Nenadic, and B. A. Bauer, "The feasibility of wearable and self-report stress detection measures in a

semi-controlled lab environment," *IEEE Access*, vol. 9, pp. 102 053–102 068, 2021.

[169] M. Nazeer, S. Salagrama, P. Kumar, K. Sharma, D. Parashar, M. Qayyum, and G. Patil, "Improved method for stress detection using bio-sensor technology and machine learning algorithms," *MethodsX*, vol. 12, p. 102581, 2024.

[170] A. Saeed and S. Trajanovski, "Personalized driver stress detection with multi-task neural networks using physiological signals," arXiv preprint arXiv:1711.06116, 2017. [Online]. Available: https://arxiv.org/abs/1711.06116

[171] A. Saeed, T. Ozcelebi, J. Lukkien, J. B. F. van Erp, and S. Trajanovski, "Model adaptation and personalization for physiological stress detection," in *Proc. 2018 IEEE 5th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2018, pp. 209–216.

[172] A. Alkurdi, M. He, J. Cerna, J. Clore, R. Sowers, E. T. Hsiao-Wecksler, and M. E. Hernandez, "Extending anxiety detection from multimodal wearables in controlled conditions to real-world environments," *Sensors*, vol. 25, no. 4, p. 1241, 2025.

[173] N. K. Sahu, S. Gupta, and H. R. Lone, "Are anxiety detection models generalizable? a cross-activity and cross-population study using wearables," arXiv preprint arXiv:2504.03695, 2025. [Online]. Available: https://arxiv.org/abs/2504.03695

[174] V. Mishra, S. Sen, G. Chen, T. Hao, J. Rogers, C.-H. Chen, and D. Kotz, "Evaluating the reproducibility of physiological stress detection models," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 4, pp. 147:1–147:27, Dec. 2020.

[175] A. Arsalan and M. Majid, "Human stress classification during public speaking using physiological signals," *Comput. Biol. Med.*, vol. 133, p. 104377, Jun. 2021.

[176] C. Marzi, A. Greco, E. P. Scilingo, and N. Vanello, "Towards a model of arousal change after affective word pronunciation based on electrodermal activity and speech analysis," *Biomed. Signal Process. Control*, vol. 67, p. 102517, 2021.

[177] A. Nechyporenko, M. Frohme, Y. Strelchuk, V. Omelchenko, V. Gargin, L. Ishchenko, and V. Alekseeva, "Galvanic skin response and photoplethysmography for stress recognition using machine learning and wearable sensors," *Appl. Sci.*, vol. 14, no. 24, p. 11997, 2024.

[178] W. Umer, "Simultaneous monitoring of physical and mental stress for construction tasks using physiological measures," *J. Build. Eng.*, vol. 46, p. 103777, 2022.

[179] G. Vos, K. Trinh, Z. Sarnyai, and M. R. Azghadi, "Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices," *J. Biomed. Inform.*, vol. 148, p. 104556, 2023.

[180] H. V. Skat-Rørdam, M. H. Knudsen, S. N. Knudsen, N. N. Lønfeldt, S. Das, and L. K. H. Clemmensen, "Applying pre-trained deep-learning model on Wrist Angel data – an analysis plan," arXiv preprint arXiv:2312.09052, 2023. [Online]. Available: https://arxiv.org/abs/2312.09052

[181] F. D. Martino and F. Delmastro, "High-resolution physiological stress prediction models based on ensemble learning and recurrent neural networks," in *Proc. 2020 IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2020, pp. 1–6.

[182] P. Siirtola and J. Röning, "Comparison of regression and classification models for user-independent and personal stress detection," *Sensors*, vol. 20, no. 16, p. 4402, 2020.

[183] L. Furia, M. Tortora, P. Soda, and R. Sicilia, "Exploring early stress detection from multimodal time series with deep reinforcement learning," in *Proc. 2023 IEEE Int. Conf. Bioinform. Biomed. (BIBM)*, 2023, pp. 1917–1920.

[184] Y. Han, P. Zhang, M. Park, and U. Lee, "Systematic evaluation of personalized deep learning models for affect recognition," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, no. 4, pp. 206:1–206:35, Nov. 2024.

[185] H. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, and W. Zuo, "Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation," arXiv preprint arXiv:1705.00609, 2017. [Online]. Available: https://arxiv.org/abs/1705.00609

[186] A. Singh, "CLDA: Contrastive learning for semi-supervised domain adaptation," arXiv preprint arXiv:2107.00085, 2021. [Online]. Available: https://arxiv.org/abs/2107.00085

[187] K. Oh, H. Park, G. H. Lee, and J. K. Choi, "Supervised contrastive learning-based stress detection for wearable sensor-based healthcare applications," *Future Gener. Comput. Syst.*, vol. 175, p. 108058, 2026.

[188] A. Bauer, S. Trapp, M. Stenger, R. Leppich, S. Kounev, M. Leznik, K. Chard, and I. Foster, "Comprehensive exploration of synthetic data generation: A survey," arXiv preprint arXiv:2401.02524, 2024. [Online]. Available: https://arxiv.org/abs/2401.02524

[189] M. Ehrhart, B. Resch, C. Havas, and D. Niederseer, "A conditional GAN for generating time series data for stress detection in wearable physiological sensor data," *Sensors*, vol. 22, no. 16, p. 5969, 2022.

[190] M. Paniagua-Gómez and M. Fernandez-Carmona, "Trends and challenges in real-time stress detection and modulation: The role of the IoT and artificial intelligence," *Electronics*, vol. 14, no. 13, p. 2581, 2025.

[191] R. K. Nath, H. Thapliyal, and A. Caban-Holt, "Machine learning based stress monitoring in older adults using wearable sensors and cortisol as stress biomarker," *J. Signal Process. Syst.*, vol. 94, no. 6, pp. 513–525, Jun. 2022.

[192] A. Saeed, V. Ungureanu, and B. Gfeller, "Sense and learn: Self-supervision for omnipresent sensors," arXiv preprint arXiv:2009.13233, 2021. [Online]. Available: https://arxiv.org/abs/2009.13233

[193] T. Islam and P. Washington, "Personalized prediction of recurrent stress events using self-supervised learning on multimodal time-series data," arXiv preprint arXiv:2307.03337, 2023. [Online]. Available: https://arxiv.org/abs/2307.03337

[194] ——, "Individualized stress mobile sensing using self-supervised pre-training," *Appl. Sci.*, vol. 13, no. 21, p. 12035, 2023.

[195] K. Matton, R. Lewis, J. Guttag, and R. Picard, "Contrastive learning of electrodermal activity representations for stress detection," in *Proc. Conf. Health, Inference, Learn.*, ser. Proceedings of Machine Learning Research, B. J. Mortazavi, T. Sarker, A. Beam, and J. C. Ho, Eds., vol. 209. PMLR, 2023, pp. 410–426.

[196] A. Almadhor, G. A. Sampedro, M. Abisado, S. Abbas, Y.-J. Kim, M. A. Khan, J. Baili, and J.-H. Cha, "Wrist-based electrodermal activity monitoring for stress detection using federated learning," *Sensors*, vol. 23, no. 8, p. 3984, 2023.

[197] B. C. Gül, S. Nadig, S. Tziampazis, N. Jazdi, and M. Weyrich, "FedMultiEmo: Real-time emotion recognition via multimodal federated learning," arXiv preprint arXiv:2507.15470, 2025. [Online]. Available: https://arxiv.org/abs/2507.15470

[198] M. A. Fauzi, B. Yang, and B. Blobel, "Comparative analysis between individual, centralized, and federated learning for smartwatch based stress detection," *J. Pers. Med.*, vol. 12, no. 10, p. 1584, 2022.

[199] A. B. S, P. P. K, and M. Rao, "MC-QDSNN: Quantized deep evolutionary SNN with multidendritic compartment neurons for stress detection using physiological signals," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 44, no. 4, pp. 1313–1325, Apr. 2025.

[200] O. B. Amin, V. Mishra, T. M. Tapera, R. Volpe, and A. Sathyanarayana, "Extending stress detection reproducibility to consumer wearable sensors," arXiv preprint arXiv:2505.05694, 2025. [Online]. Available: https://arxiv.org/abs/2505.05694

[201] M. K. Moser, M. Ehrhart, and B. Resch, "An explainable deep learning approach for stress detection in wearable sensor measurements," *Sensors*, vol. 24, no. 16, p. 5085, 2024.

[202] Shikha, D. Sethia, and S. Indu, "Optimization of wearable biosensor data for stress classification using machine learning and explainable AI," *IEEE Access*, vol. 12, pp. 169 310–169 327, 2024.

[203] N. Mandal, "Interpretable multi-task PINN for emotion recognition and EDA prediction," arXiv preprint arXiv:2505.18169, 2025. [Online]. Available: https://arxiv.org/abs/2505.18169