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

Abstract—**Electrodermal Activity (EDA) has emerged as a prominent modality for pervasive stress monitoring due to its high sensitivity to sympathetic arousal and computational efficiency. However, research progress is currently impeded by systemic fragmentation, including scattered datasets, inconsistent preprocessing pipelines, and a lack of rigorous benchmarks. Critically, existing reviews dedicated to EDA often lack the systematic depth and technical detail required to address these methodological inconsistencies. Bridging this gap, this paper presents an end-to-end systematic survey and large-scale empirical benchmark for EDA-based stress detection. First, we thoroughly review the technical pipeline from signal decomposition to modeling, while compiling a centralized resource of 21 public datasets and systematically analyzing 19 representative unimodal studies to foster reproducibility. Then, we incorporate architectural insights from advanced multimodal research and establish a standardized evaluation protocol, conducting extensive benchmarks on both traditional machine learning and state-of-the-art Deep Learning models across five diverse datasets to rigorously assess multi-domain robustness and cross-scenario generalization. Finally, we critically examine key implementation challenges, including deployment efficiency and data scalability, and provide a strategic outlook for future advancements.**

Index Terms—**Electrodermal Activity, stress detection, affective computing, machine learning, deep learning.**

I. Introduction

In modern society, mental health has become a critical determinant of overall well-being. However, the prevalence of chronic stress—exacerbated by information overload and fast-paced lifestyles—poses a severe threat to public health. While moderate ”eustress” can enhance performance, persistent ”distress” triggers prolonged activation of the sympathetic nervous system (SNS) [1], [2]. This maladaptive response leads to elevated cortisol levels and is linked to cardiovascular diseases, immune suppression, and cognitive decline [3]. With the World Health Organization projecting over one billion people will suffer from mental health issues by 2025, and remote work amplifying these risks, there is an urgent need for effective, real-time stress monitoring strategies [4], [5].

Timely stress detection enables early interventions, such as mindfulness or workload adjustment, preventing the escalation of psychological distress [6]. Traditional assessments relying on subjective questionnaires or invasive biomarkers suffer from bias and lack real-time capabilities [7], [8]. Consequently, automated detection via physiological signals has become the mainstream research direction [9]. While signals like ECG (Heart Rate Variability) [10], EEG (cortical activity) [11], and PPG (pulse waves) [12] offer valuable insights, multimodal fusion often introduces challenges in data synchronization, computational complexity, and user discomfort due to multiple sensors [13].

In contrast, unimodal Electrodermal Activity (EDA) presents a pragmatic and efficient alternative for pervasive stress monitoring [14]. EDA reflects sweat gland activity directly regulated by the SNS, making it highly sensitive to emotional and sympathetic arousal without the interference of parasympathetic activity [15]. Rather than pursuing marginal accuracy gains at the cost of system complexity, unimodal EDA offers an optimal trade-off: it provides a sufficiently reliable proxy for stress levels (0.7–0.9 correlation in standard protocols) while significantly reducing hardware costs and computational overhead compared to multimodal setups [16]–[19]. Furthermore, its non-invasive nature facilitates seamless edge deployment on wrist-worn devices. Driven by these attributes, the utility of EDA has been validated across a wide spectrum of real-world applications, ranging from safety-critical monitoring in driving and construction [20]–[22] and work state recognition [23], to cognitive state assessment for academic performance [24]–[29] and human-computer interaction (HCI) optimization [30]–[32]. These scenarios highlight EDA’s versatility and its favorable trade-off between computational efficiency and detection performance. [33].

Despite these advancements, the field is currently impeded by systemic fragmentation. First, the isolation of datasets poses a fundamental barrier; researchers typically rely on scattered, limited databases, which leaves the critical question of cross-scenario generalization largely unverified. Second, the lack of standardized processing pipelines—characterized by inconsistent signal preprocessing and feature engineering methods—renders fair benchmarking across different studies nearly impossible. Critically, existing literature reviews fail to adequately address these gaps. Surveys dedicated solely to EDA often lack comprehensive coverage and technical depth, tending to be descriptive rather than analytical regarding methodological inconsistencies. Conversely, broader stress detection surveys predominantly prioritize multimodal fusion strategies, often neglecting the specific nuances and decomposition algorithms essential to the unimodal EDA pipeline. Most importantly, both categories of reviews lack a rigorous, empirical benchmark to validate the comparative performance of state-of-the-art models. To bridge these gaps, this paper makes the following contributions:

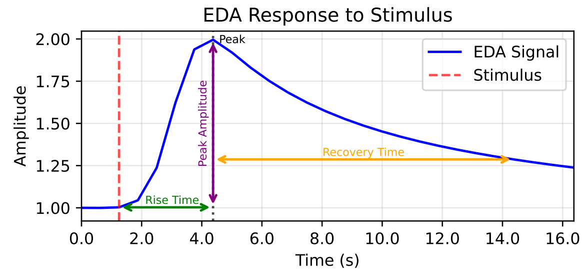
- **Comprehensive Methodology Integration & Resource Compilation:** We present first end-to-end systematic review of the unimodal EDA pipeline, from preprocessing to modeling. Concurrently, we compile a centralized resource of 21 public datasets and systematically categorize 19 representative unimodal studies to foster community reproducibility.
- **Cross-Scenario Empirical Benchmark:** We establish a standardized evaluation protocol and conduct a large-scale benchmark across five diverse datasets. Specifically, we design rigorous multi-domain fusion and cross-domain transfer experiments to evaluate both traditional machine learning and state-of-the-art Deep Learning models, providing the first empirical reference for model selection and generalization.
- **Strategic Insights & Future Outlook:** We synthesize architectural insights from advanced multimodal research to inform and optimize unimodal modeling strategies. Furthermore, we critically examine key implementation hurdles, including data scarcity, labeling quality, and practical deployment challenges, and outline a strategic roadmap for future advancements.

The remainder of this paper is organized as follows: Section II provides the background on EDA signals and presents a systematic comparison with existing surveys, alongside a comprehensive catalog of public datasets. Section III critically reviews the end-to-end technical pipeline, covering signal preprocessing, decomposition algorithms, and modeling strategies from both unimodal and multimodal perspectives. Section IV establishes the empirical benchmark, detailing the experimental protocol and analyzing the performance of ML and DL models across multi-domain fusion and cross-scenario generalization tasks. Section V discusses future research directions and challenges. Finally, Section VI concludes the paper.

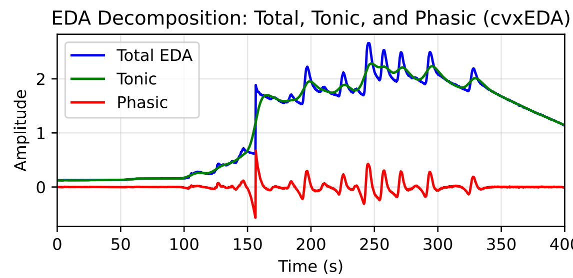
II. Background and Related Work

To identify EDA-based stress-detection studies, we searched IEEE Xplore, PubMed, Web of Science, and Google Scholar using combinations of the keywords "electrodermal activity", "skin conductance", "Galvanic Skin Response (GSR)", "stress detection", "stress recognition", "mental workload", and "wearable". We focused on human-subject studies published between 2010 and 2025 that (i) recorded EDA signals, (ii) included explicit stress labels, and (iii) reported quantitative performance of a computational model. Based on these criteria, we compiled the public datasets summarized in Table II and 19 representative unimodal EDA stress-detection works in Table III. For multimodal systems that include EDA among other signals, we do not aim for exhaustive coverage; instead, we highlight representative studies that illustrate modeling patterns relevant to EDA-only architectures.

A. EDA



(a) An illustration of EDA response



(b) An illustration of EDA decomposition

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

EDA, also known as 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 (standard sites), or the wrist (common in wearable devices). 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], [42].

As illustrated in Figure 1b, EDA can be decomposed into two core components: tonic and phasic. The tonic component, or Skin Conductance Level (SCL), represents slow fluctuations (0.05–0.5 Hz) reflecting background arousal, which is susceptible to environmental factors or individual baselines. In contrast, the phasic component, or Skin Conductance Response (SCR), consists of rapid, transient peaks (0.5–2 Hz) triggered by sudden SNS activation, such as confronting a stressor. While the superposition of multiple SCRs

Ref./Year	# Refs.	Signal Modality	Core Focus	Main Contributions	Main Limitations
Part I: EDA-Specific Surveys					
1. [34]/2019	14	EDA Only	EDA for Stress Detection	1. The first survey explicitly dedicated to EDA-based stress detection.	1. Limited literature coverage (only 14 articles). 2. Outdated (published in 2019), missing recent advancements. 3. Primarily descriptive listing of works, lacking deep insight or analysis.
2. [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).
3. [36]/2020	103	EDA Only	EDA with ML for Stress Detection	1. 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).
4. [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.
5. [38]/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.
Part II: Multimodal Surveys (Physiological & General)					
6. [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. 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.
7. [39]/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.
8. [40]/2025	202	Multimodal Data (Physiological, Behavioral, Environmental)	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.
9. [41]/2025	185	Multimodal Data (Physiological, Speech, Facial, Social Media)	Application of DL in Multimodal Stress Detection	1. Comprehensively overviews the application of DL in multimodal stress detection, highlighting its advantage in automatic feature extraction. 2. Details the methods, performance, and public datasets for each modality.	1. Focuses exclusively on DL, neglecting traditional ML approaches. 2. The specific discussion and depth regarding EDA are relatively limited.
Proposed Work					
This Study	213	Primarily EDA	A Comprehensive Survey and Empirical Benchmark	1. **End-to-End Review & Resources**: Presents the first systematic survey of the unimodal EDA pipeline and compiles 21 public datasets and 19 unimodal studies to foster reproducibility. 2. **Cross-Scenario Benchmark**: Establishes a standardized protocol to benchmark ML and DL models across multi-domain fusion and cross-domain generalization tasks. 3. **Strategic Outlook**: Synthesizes architectural insights from advanced multimodal research to optimize unimodal strategies and map future advancements.	1. The current benchmark focuses on supervised learning; advanced paradigms such as self-supervised learning and domain adaptation are discussed as future directions rather than empirically tested.

TABLE I: Comparison of Related Surveys. The table is categorized into two sections: **Part I** reviews surveys explicitly targeting EDA, while **Part II** covers high-impact surveys on broader multimodal stress detection. The column **'# Refs.'** indicates the number of papers reviewed in each survey.

creates a complex signal morphology, their amplitude and frequency directly encode the intensity of external stimuli. EDA is particularly effective for stress detection because stress inherently triggers the SNS-dominated "fight-or-flight" response, causing a quantifiable surge in sweat gland activity. Section III-B provides a detailed discussion of decomposition algorithms.

B. Comparison with other surveys

As summarized in Table I, we have systematically reviewed existing surveys closely related to the theme of this study. To ensure the quality and relevance of the comparison, we selected representative surveys from two core categories: those explicitly targeting EDA as the primary research object, and high-impact surveys concentrating on the broader stress detection scenario. The "Cited Articles" metric in the table reflects the breadth of literature covered, while "Focus" clarifies the core area of attention. Existing research can be broadly classified into two categories: EDA-centric surveys and surveys centered on multimodal stress detection.

1) EDA-centric surveys: Early works, such as the study by [35], pioneered the focus on EDA signal acquisition and processing techniques but did not extend to downstream modeling and applications. [34] was the first survey to explicitly target EDA-based stress detection; however, as an early work, it covered a limited number of papers and primarily listed existing works without providing deep methodological insights or analysis of recent advancements. Subsequent works began to cover the technical pipeline but were often constrained by narrow application scenarios, such as focusing solely on emotion recognition in educational contexts [37] or specific care environments [38], limiting their generalizability. Among these, [36] is the closest in scope to our study. However, it conflates the distinct fields of "stress detection" and "emotion recognition" and lacks a systematic benchmark.

2) surveys centered on multimodal stress detection: Compared to unimodal EDA works, research in this category is generally more mature and systematic. These surveys typically review the field from the perspectives of signal input, feature extraction, and classification models (covering both traditional ML and feature-based DL) [43]–[46]. Some works extend their scope to include specialized modalities, such as biochemical sensors (e.g., salivary cortisol) [46], contactless technologies like spectral imaging and rPPG [47], and even behavioral, environmental (e.g., CO2 concentration), or social media data [40], [41], [44]. Specific contexts, such as knowledge work environments [40] and student populations [48], have also been discussed.

While these multimodal surveys comprehensively address common issues—such as the "lab-to-wild" gap, device heterogeneity, and the risk of overfitting in small datasets—they face a trade-off. Due to the sheer breadth of modalities covered, they often lack depth in discussing the specific technical details of the EDA pipeline. Few surveys delve into critical preprocessing parameters (e.g., filter cutoff frequencies, downsampling rates), specific artifact removal strategies, or the comparative effectiveness of decomposition algorithms [9], [39]. This lack of detail can lead to ambiguity in the research pipeline, hindering the standardization of EDA-based methods. Consequently, despite the frequent acknowledgment of EDA's importance in these surveys [44], [46], there remains a significant gap for a survey that offers both a comprehensive, end-to-end review of the unimodal EDA pipeline and a rigorous empirical benchmark.

C. Datasets

To promote reproducible research and enhance model generalizability, 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 relevant to stress detection. While previous surveys have touched upon datasets in affective computing [49]–[52], they often fail to fully meet the specific needs of EDA-based stress detection research. For instance, [49] primarily focuses on cognitive load rather than the broader context of psychological stress; [50] is limited in scope, covering only 9 datasets with just 5 involving EDA; and while [51], [52] comprehensively summarize numerous stress datasets, they do not specifically target scenarios where EDA is the core modality. To address this gap, this study catalogs the key attributes of each dataset—including subject count, experiment duration, signal modalities, collection devices/frequencies, experimental paradigms, and access links—serving as a comprehensive reference guide for researchers. However, despite their public availability, practical application remains fraught with several challenges, as these factors collectively constitute significant practical hurdles for researchers utilizing public data resources:

- **Data accessibility barriers:** Some datasets (e.g., VeroBIO [61], EmpathicSchool [62], StressID [65], K-EmoPhone [68], ForDigitStress [69]) are not available for direct download, requiring researchers to submit application forms or contact authors via email, which increases the complexity and time cost of acquisition.
- **Technical usability barriers:** Early datasets (e.g., RW-Driving [53]) are often limited by outdated acquisition devices and protocols, while others (e.g., SWELL [54]) utilize proprietary formats (e.g., S00) that require specific software for parsing, adding to the preprocessing burden.
- **Uncertain release status:** The release status of several promising datasets remains uncertain. Highly valuable resources—such as CIRVR [74] (autistic youth interviews), LAUREATE [75] (longitudinal ecological study), MMSD [76] (multimodal with cortisol), and MuSE [77] (real exam stressors)—have been detailed

Dataset Name	Year	Subj.	Dur.	Signals	Device/Frequency	Description	Link
1. RWDriving [53]	2008	27	50–90 min	EDA, EKG, EMG, RESP	FlexComp/31 Hz	Real-world driving scenario	Link
2. SWELL [54]	2014	25	138 min	EDA, HRV, ECG	Mobi device (TMSI)/2048 Hz	Knowledge work, stress induced by email interruptions	Link
3. UTD [55]	2017	20	37 min	EDA, ACC, Temp, SpO2, HR	Affectiva/8 Hz	Physical, cognitive, and emotional stressors	Link
4. WESAD [56]	2018	15	120 min	EDA, ACC, BVP, HR, IBI, TEMP	RespiBAN/700 Hz, Empatica E4/4 Hz	Baseline, stress, amusement, and meditation states	Link
5. AffectiveROAD [57]	2018	13	120 min	EDA, ACC, BVP, HR, IBI, TEMP	Empatica E4/4 Hz	Real-world driving scenario	Link
6. CLAS [58]	2019	62	30 min	EDA, ECG, PPG, ACC	Shimmer GSR Unit/256 Hz	Mathematical, cognitive, and emotional stressors	Link
7. Toadstool [59]	2020	10	60 min	EDA, ACC, BVP, IBI, HR, TEMP	Empatica E4/4 Hz	Playing the game "Super Mario Bros."	Link
8. MAUS [60]	2021	22	35 min	EDA, ECG, PPG (Finger/Wrist), IBI	Procomp Infiniti/256 Hz	N-back cognitive task	Link
9. VerBIO [61]	2022	55	NA	EDA, ECG, BVP, TEMP, ACC	Empatica E4/4 Hz	Public speaking anxiety	Link
10. EmpathicSchool [62]	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 [63]	2022	10	4.5 hrs	EDA, HR, BVP, TEMP, IBI, ACC	Empatica E4/4 Hz	Real-world exam scenario	Link
12. Nurse Stress Prediction [64]	2022	15	1 week	EDA, HR, TEMP, ACC, IBI, BVP	Empatica E4/4 Hz	Hospital work scenario	Link
13. StressID [65]	2023	65	35 min	EDA, ECG, RESP	Bio SignalsPlux system/500 Hz	11 tasks including cognitive, emotional, and speech	Link
14. UBFC-Phys [66]	2023	56	9 min	EDA, BVP, RPPG, PRV, TEMP	Empatica E4/4 Hz	Speech and mental arithmetic tasks	Link
15. BIOSTRESS [67]	2023	28	90 min	EDA, ACC, TEMP, BVP, HR, IBI	Empatica E4/4 Hz	Stressful interview and final exam scenarios	Link
16. K-EmoPhone [68]	2023	77	7 days	EDA, ACC, HR, TEMP, RRI	Microsoft Band 2/5 Hz	Seven days of daily life monitoring	Link
17. ForDigitStress [69]	2023	40	75 min	EDA, PPG, Voice	IOM biofeedback sensor/5 Hz	Job interview scenario	Link
18. MUMTG [70]	2024	20	NA	EDA, BVP, HR, IBI, TEMP, Pupil	Empatica E4/4 Hz	Building blocks task	Link
19. EmpaticaE4Stress [71]	2024	29	40 min	EDA, ACC, BVP, HR, IBI, TEMP	Empatica E4/4 Hz	Simulated work scenario	Link
20. WorkStress3D [72]	2024	20	8.75 hrs	EDA, BVP, TEMP, ACC	Empatica E4 Wristband/4 Hz	In-office work day	Link
21. EmoPairCompete [73]	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. **Abbreviations**: Accelerometer (ACC), Blood Volume Pulse (BVP), Electrocardiogram (ECG), Electrodermal Activity (EDA), Electromyogram (EMG), Heart Rate (HR), Heart Rate Variability (HRV), Inter-Beat Interval (IBI), Photoplethysmogram (PPG), Pulse Rate Variability (PRV), Remote Photoplethysmography (RPPG), Respiration (RESP), R-R Interval (RRI), Blood Oxygen Saturation (SpO2), Temperature (TEMP).

in literature with intentions for open-sourcing. Yet, at the time of this survey, they remain unreleased.

III. Technical Review of EDA Pipeline

In the early 2010s, a series of pioneering studies laid the methodological foundation for automated stress detection using wearable EDA sensors. Setz et al. [78] demonstrated that unimodal EDA could distinguish psychological stress from cognitive load with an accuracy of up to 82.8%, establishing strong evidence for its feasibility. Subsequently, [79] extended this validation to real-world settings, exploring automated stress identification in daily work scenarios. Concurrently, critical challenges in the field were identified, and attempts were made to address them. In the context of multimodal fusion, Choi et al. [80] effectively distinguished physiological states via unsupervised clustering by integrating EDA with heart rate, respiration, and EMG. Furthermore, Kurniawan et al. [81] highlighted the robustness of EDA, confirming its superior performance and generalization over speech signals under user-independent conditions. Addressing individual variability, [82] pioneered personalized adaptive models by optimizing SVM loss functions based on electrodermal reactivity. Parallel to these algorithmic advances, limitations in early technology necessitated significant research into hardware innovations for robust data acquisition [83]–[85].

A. Artifact Removal

Motion artifacts, stemming from skin deformation and muscle activity, are primary sources of interference in EDA signals. Strategies to mitigate these artifacts fall into two categories: *signal correction* (suppression) and *quality assessment* (segment rejection).

1) Signal correction: Traditional non-adaptive methods, such as exponential smoothing and low-pass filtering, often fail to handle high-amplitude artifacts and risk distorting valid signals [86], [87]. To address this, adaptive denoising frameworks have been developed. While Chen et al. introduced a stationary wavelet transform approach, it proved sensitive to parameter tuning [88]. Subsequent improvements include modeling wavelet coefficients with a Laplace distribution to replace computationally expensive Gaussian Mixture Model [89], and utilizing biophysical models via extended Kalman filters for robust noise suppression [90]. In specific contexts like driving, dual-hand fusion strategies have successfully filtered sharp spikes caused by steering maneuvers [91], [92].

2) artifact identification: Early supervised methods used SVMs to classify artifacts but relied heavily on manual annotation [93]. Zhang et al. argued that supervised approaches scale poorly and demonstrated the efficacy of unsupervised learning for this task [94]. Conversely, to avoid the "black-box" nature of ML, Kleckner et al. proposed a transparent, rule-based method using heuristic physiological constraints, enhancing the algorithm's interpretability [95].

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 [96]. Subsequently, they enhanced the model's time-invariance by introducing physiological constraints [97] and used Dynamic Causal Modeling (DCM) to simulate sudomotor nerve activity in anticipatory responses [98]. Concurrently, Benedek et al. developed a technique based on non-negative deconvolution, ensuring the non-negativity of the driver signal through Gaussian elimination [99], [100]. These pioneering model-driven methods laid the foundation for subsequent research [101]–[103]. 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 [104]. Bach et al. employed a matching pursuit algorithm to capture sparse, spontaneous SCR events to infer tonic sympathetic arousal levels [105]. Among these, the cvxEDA method, developed by Greco et al. and based on convex optimization, became a significant milestone [106]. 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 [107], compressed sensing [108], a non-negative sparse deconvolution algorithm based on an overcomplete dictionary (sparsEDA) [109], as well as Bayesian inference models based on state-space physiological representations [110]–[112] and a multichannel decomposition framework [113].

Although numerous decomposition algorithms have been proposed, their availability as accessible open-source software remains uneven. Widely used tools such as Ledalab (including CDA and DDA), SCRalyze, and PsPM are primarily distributed as standalone applications or MATLAB-dependent toolboxes. This dependence can limit flexibility and hinder integration into cross-platform, automated analysis workflows [114]–[116]. In contrast, algorithms like cvxEDA and sparsEDA have been released as independent open-source code libraries on GitHub, a shift that has greatly facilitated their widespread adoption and dissemination within the research community [117], [118].

Empirical studies comparing these methods in practical scenarios reveal that optimal performance is often context-dependent [33], [119]–[121]. For instance, [120] found sparsEDA to be superior in computational efficiency and classification accuracy for stress detection, whereas [33] reported that CDA slightly outperformed cvxEDA in cognitive tasks like the Stroop test. Meanwhile, [121] identified TVSymp as the most effective for emotion recognition among six algorithms. Despite these variations, a critical consensus exists: the performance difference among mainstream algorithms (e.g., CDA, cvxEDA, sparsEDA) in final classification tasks is typically marginal, falling within a 1–2% range. This implies that while decomposition is essential, differences in final application performance stem more from downstream feature extraction and modeling strategies than from the superiority of the decomposition algorithm itself.

Fortunately, the emergence of comprehensive physiological signal processing libraries such as NeuroKit2 [122] and BioSPPy [123] has bridged the gap between advanced algorithms and practical application. As open-source Python toolkits, they integrate core decomposition functions like cvxEDA and sparsEDA into standardized, easy-to-call APIs, significantly lowering the technical barrier for researchers.

C. Nonlinear Methods

Beyond standard time- and frequency-domain metrics, nonlinear dynamics offer deeper insights into the intrinsic correlation between EDA and stress. Early research prioritized frequency-specific analysis: POSADA-QUINTERO et al. introduced EDASymp, measuring normalized power in the 0.045–0.25 Hz band [124], and subsequently proposed the more sensitive TVSymp, which utilizes time-varying complex demodulation (0.08–0.24 Hz) to track transient sympathetic tone [107]. Alternatively, focusing on computational efficiency, [125] demonstrated that simple regression-based features—specifically the slope and residual area within short windows—can effectively distinguish relaxation states.

Recent trends have shifted towards complexity science to quantify signal dynamics. [126] proposed ComEDA, a metric that quantifies dynamic spatial complexity by applying quadratic Rényi entropy to angular distances in a reconstructed phase space, showing superior performance in differentiating stress levels. To capture dynamics across temporal resolutions, [127] extended this to MComEDA. By integrating coarse-graining processes and calculating the area under the multi-scale complexity curve, MComEDA outperforms both EDASymp and single-scale ComEDA in distinguishing emotional arousal and valence.

D. Unimodal

To rigorously assess the state-of-the-art in unimodal EDA stress detection, we conducted a systematic review of representative studies, as summarized in Table III. This compilation, which details datasets, preprocessing pipelines, feature engineering, and validation protocols across traditional ML and emerging DL approaches, highlights pivotal trends and systemic challenges. In dataset selection, the field has matured from relying on self-collected data to utilizing standardized public datasets like WESAD, significantly enhancing reproducibility. Conversely, methodology remains fragmented; the lack of standardized preprocessing and feature engineering protocols continues to impede fair cross-study benchmarking. In modeling, Support Vector Machines (SVM) and ensemble methods (e.g., XGBoost) dominate the ML landscape due to their competitiveness. Crucially, regarding evaluation standards, Leave-One-Subject-Out (LOSO) cross-validation and strict user-independent testing are establishing themselves as the gold standard, reflecting a growing emphasis on generalization against physiological inter-subject variability.

Within the ML domain, several studies distinguish themselves through innovative problem-solving. Addressing individual heterogeneity, [132] proposed a "Localized Supervised Learning" framework that clusters physiological response patterns to train specialized classifiers, significantly outperforming global models. Focusing on architecture and temporal resolution, [134] validated the superiority of Deep SVMs (DSVM) over traditional classifiers and identified a performance saturation point at a 5-second window length, noting that window overlap had negligible impact. The efficacy of unimodal EDA is further corroborated by [16] and [17], which demonstrated across multiple datasets that pure EDA-based classification frequently surpasses other single modalities and even multimodal combinations. Furthermore, to bridge the gap to real-time application, [139] formalized the trade-off between latency and accuracy by introducing the "Time to Detection" (TTD) metric: $\mathrm{TTD} = t_{\text{length}} + (\omega - 1)t_{\text{step}}$, where t_{length} is the window length, t_{step} is the step size, and ω represents the consecutive segments required for a decision.

In the DL domain, end-to-end methodologies are increasingly adopted to circumvent manual feature engineering, targeting specific implementation bottlenecks. To mitigate data scarcity, [141] leveraged biophysical models to synthesize high-fidelity EDA data under varying stress levels, effectively augmenting training sets. Addressing computational constraints on edge devices, [142] employed Neural Architecture Search (NAS) to engineer TinyStressNet, a lightweight CNN that reduces model size and complexity while maintaining performance parity with large-scale benchmark models, demonstrating great potential for real-time stress monitoring on resource-constrained edge devices.

Beyond stress detection, methodologies from EDA-based emotion recognition offer valuable cross-domain insights. [42] systematically evaluated 40 features, highlighting the efficacy of acoustic-domain features like Mel-Frequency Cepstral Coefficients (MFCCs), thereby suggesting new avenues for feature engineering. Additionally, [143] pioneered the application of Graph Signal Processing (GSP) to EDA, demonstrating that transforming time-series signals into graph structures captures complex nonlinear patterns more effectively, providing a novel theoretical perspective for future model

Ref.	Dataset(s)	Preprocessing	Features Info. (Domain) (Selection, Count)	Model(s)	Validation	Reported Performance
Part I: Machine Learning (ML) Approaches						
1. [128]	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)
2. [129]	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)
3. [130]	RWDriving	Low-pass filter, 5 min window	18 (Time) (Fisher projection, 2)	LDA	LOOCV	81.82% Acc (3-class stress level)
4. [131]	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
5. [132]	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)
6. [133]	WESAD	20 Hz downsample, 1s Moving Average Filter, Min-Max Norm	No specific Number (Statistical features, DL features)(None)	**k-NN**, SVM, RF, NB	10-fold CV	91.6% Acc For DL features, 90 % For Statistical features (2-class)
7. [134]	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)
8. [135]	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)
9. [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
10. [136]	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
11. [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
12. [137]	WESAD	8Hz upsample, 60s window, 0.25s stride, cvxEDA decomp.	87 (Time) (Forward selection, 5)	**AdaBoost**, RF, SVM	LOSO	97.03% Acc
13. [138]	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
14. [139]	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
15. [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
16. [140]	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. [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
Part II: Deep Learning (DL) Approaches						
18. [141]	Self-collected; physical activity	2min window (1min overlap)	End-to-End	FCN	Train on synthetic, test on real	96% (Synthetic) / 84% (Real) Acc
19. [142]	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 table is divided into Part I (Machine Learning) and Part II (Deep Learning). The best-performing model in each study is indicated in **bold**. **Features Column Explanation**: Data is presented as: *Total Number of Extracted Features (Feature Domains) (Feature Selection Method, Final Selected Count)*. **Abbreviations**: *Models*: CNN (Convolutional Neural Network), 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), LR (Logistic Regression), NAS (Neural Architecture Search), NB (Naïve Bayes), RF (Random Forest), SEL (Stacking Ensemble Learning), SVM (Support Vector Machine). *Validation*: CV (Cross-Validation), LOOCV (Leave-One-Out Cross-Validation), LOSO (Leave-One-Subject-Out). *Others*: MIC (Maximal Information Coefficient), TSST (Trier Social Stress Test).

innovation.

E. Insights from Multimodal Methods

While this survey prioritizes unimodal EDA efficiency, multimodal research offers critical theoretical and architectural insights relevant to EDA-specific modeling. Multimodal frameworks typically augment EDA with physiological indicators—such as ECG/PPG-derived Heart Rate Variability (HRV) [91], [144]–[154], Skin Temperature (ST) [147]–[149], [151], EMG [145], [152], Respiration (RSP) [145], [152], [155], and EEG [145], [155]—to capture a broader stress response spectrum. The evolution of these methodologies provides a roadmap for advancing unimodal EDA modeling across three key dimensions.

1) Architectural Evolution: Crucially, multimodal architectures offer specific design patterns and evolutionary pathways directly applicable to unimodal EDA analysis. The field is fundamentally shifting from manual feature engineering towards autonomous representation learning. Traditional ML approaches predominantly follow a "feature extraction-fusion-classification" pipeline, which relies heavily on manual engineering and often treats EDA crudely without component separation [148]–[151], [156]. While hybrid DL approaches (handcrafted features + NN) exist [154], [155], [157]–[159], they incur computational costs and suffer from "black-box" interpretability issues. In contrast, end-to-end approaches that ingest raw or minimally preprocessed signals leverage DL to autonomously learn hierarchical representations, demonstrating superior feasibility in capturing complex physiological dynamics [160], [161].

In terms of specific structural design, research indicates that early feature interaction within convolutional layers outperforms late decision-level fusion [162]. This motivates designs like the "Multimodal Transfer Module" (MMTM) for cross-modal calibration at high-level feature layers [163], hierarchical architectures with intermediate fusion [164], [165], and dual-channel frameworks separating "modality-invariant" from "modality-specific" features [166]. These architectures are reinforced by findings that ANS-based signals (ECG, EDA, RESP) share strong generalization potential distinct from motion signals [167]. Furthermore, adapting computer vision concepts—such as ResNeXt-style parallel convolutions [168], hybrid CNN-Transformer models [169], and FCN-LSTM stacks [170]—illustrates how advanced mechanisms can balance local feature extraction with global temporal modeling.

2) Advanced Signal Representation: To fully exploit the capabilities of 2D-CNNs, research has focused on transforming 1D time-series into 2D image representations. Approaches include converting signals into Continuous Recurrence Plots [171], or stacking feature maps derived from FFT, cube root transforms, and Constant Q Transforms (CQT) [172]. Similarly, direct conversion of raw signals into grayscale spectrograms has proven effective for CNN backbones [173]. Notably, encoding methods like Gramian Angular Summation Field (GASF) and Gramian Angular Difference Field (GADF) have gained traction for their computational efficiency over Markov Transition Fields (MTF), enabling 2D-CNNs to outperform traditional SVMs by encoding temporal dependencies into pixel space [174], [175]. Beyond supervised learning, unsupervised techniques, including autoencoders with pseudo-inverse learning [19] or hierarchical frequency-domain representations [176], further demonstrate pathways for optimizing feature learning without labeled data.

3) Context-Awareness and Real-World Generalization: Transitioning from laboratory to "in-the-wild" environments introduces significant challenges regarding context and individual variability. Context-based frameworks utilize auxiliary sensors (e.g., accelerometers) to differentiate physical activity from psychological stress [147], [148], [150]. However, designs must avoid methodological pitfalls such as using experimental phases as features [177] or relying on impractical synchronized self-reports [178]. Sophisticated adaptive frameworks address this by using motion data to dynamically activate context-appropriate classifier branches [179]. To handle inter-subject variability, personalized modeling has adopted Multi-Task Learning (MT-NN) [180], domain adaptation [181], demographic-aware feature engineering [145], [159], and cluster-based specialized models [152].

Systematic evaluations of the "lab-to-wild" generalization gap reveal critical insights for data strategy. Contrary to the assumption that real-world training data is always superior, models trained on high-quality laboratory data may generalize better to noisy real-world settings than those trained on noisy data directly [153], [182]. This underscores the value of clean, well-annotated datasets for initial model training. Additionally, studies highlight the need to evaluate generalization across different activities and populations [183] and address the crisis of reproducibility [184]. Research scenarios have diversified to improve ecological validity, spanning programming competitions [150], public speaking [185], speech activities [186], and simulated threats [187], alongside specific populations like those with mild cognitive impairment [146]. Notably, unimodal EDA frequently outperforms multimodal setups in both accuracy and cost-efficiency in these diverse settings, reinforcing the premise of this survey [18], [19]. Finally, data scarcity and target definition remain pivotal. Ensemble learning on merged datasets [188] and "pre-training plus fine-tuning" paradigms [189] are emerging solutions for limited data. Concurrently, the field is shifting from classification to regression tasks to predict continuous stress scores [190]. This approach preserves granular information lost in binary labeling and facilitates graded, timely interventions [191]. Underlying these high-level strategies is the critical trade-off in signal windowing: while longer windows theoretically increase information, performance saturates, necessitating a balanced selection of

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

window parameters to optimize accuracy against real-time latency [158], [172].

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.

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. Notably, we deliberately excluded explicit pre-filtering based on three strategic considerations: 1) **Inherent Robustness**: Modern decomposition algorithms, specifically cvxEDA, possess inherent noise suppression capabilities sufficient to handle artifacts such as sudden signal spikes; 2) **Signal Preservation**: Given that common devices like the Empatica E4 sample at 4 Hz, the Nyquist frequency is limited to 2 Hz. Since the stress-correlated phasic component contains frequency content up to 2 Hz, applying standard low-pass filters risks attenuating critical physiological information; and 3) **Benchmarking Objectivity**: A streamlined pipeline minimizes hyperparameter dependency, thereby establishing a more reproducible benchmark.

Accordingly, our standardized pipeline consists of the following steps:

1) **Resampling**: All EDA signals from all datasets were uniformly resampled to 4 Hz to align temporal resolution.
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. Crucially, this step was performed *after* decomposition to preserve the complete morphology of transient stress responses.

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.

- **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.
- **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. Regarding the architectural depth, we implemented the MLP and CNN models with a three-layer structure, while the recurrent networks (RNN, GRU, LSTM, BiLSTM) and the Vision Transformer (ViT) were configured with a single layer.

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. Except for SVM and DT, model performance saturates at approximately 22 features. 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.

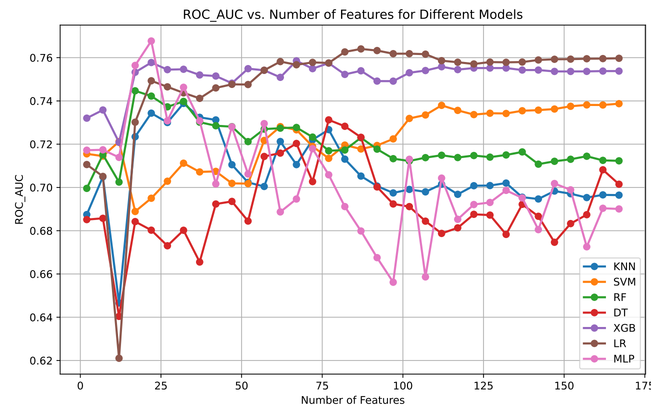


Fig. 2: ROC AUC vs. Number of Features for Different Models on the Combined Dataset Test Set.

Table VI and Table VII summarize the optimal performance of the ML and DL models on the Combined Dataset. Two critical observations emerge from these results.

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**

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

First, the best-performing DL model (CNN, 72.35%) significantly outperforms the top ML model (MLP, 66.55%). This performance gap underscores the limitations of traditional ML in handling heterogeneous data from diverse domains (e.g., driving, cognitive tasks, and social stressors), where fixed, handcrafted features fail to generalize. In contrast, the end-to-end CNN leverages hierarchical feature learning to autonomously extract robust representations from raw signals, effectively adapting to cross-scenario variability.

Crucially, the models consistently exhibit low sensitivity paired with high specificity, implying a proficiency in identifying non-stress states but a struggle to detect stress events. This asymmetry likely stems from intrinsic signal characteristics rather than simple data imbalance. The ”non-stress” state represents a stable physiological baseline, forming a compact and homogeneous cluster in the feature space. Conversely, the ”stress” state, driven by diverse SNS responses to varying stimuli (cognitive, emotional, physical), manifests as a diffuse and irregular distribution with high intra-class variance. This structural heterogeneity complicates the delineation of precise decision boundaries, presenting a fundamental challenge for future algorithmic design.

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 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

Tables IX and Table X demonstrate the severity of domain shift, where models trained on single source domains suffer significant degradation when transferred to the independent EmpaticaE4Stress dataset. This confirms that simple data aggregation in multi-domain fusion fails because learned physiological patterns remain tightly coupled to specific experimental paradigms rather than the underlying stress response.

Notably, the model trained on AffectiveROAD achieved the highest transfer performance. The CNN (AUC 72.42%) outperformed both the corresponding MLP (AUC 67.38%) and, critically, the target domain baseline (AUC 70.16%). This result reveals two notable implications.

First, **ecological validity outweighs data volume**. AffectiveROAD’s real-world driving environment generates physiological responses with higher generalizability than controlled laboratory settings, suggesting that future research should prioritize high-fidelity core datasets for pre-training over indiscriminate aggregation.

Second, **DL representations exhibit superior transferability**. The CNN’s ability to surpass the baseline implies it extracted abstract representations closer to the intrinsic physiological stress response, which proved more robust than patterns learned directly from the target domain’s limited scenario. Conversely, the MLP was constrained by its fixed 22-feature set, lacking the adaptability required for effective cross-domain transfer.

These benchmarks systematically quantify the advantages of DL for unimodal EDA. Despite the nascent state of end-to-end DL in this subfield (see Table III), these empirical results provide compelling evidence for a paradigm shift from traditional ”feature engineering + ML” to ”end-to-end DL.” This study establishes a rigorous performance baseline to guide future advancements in model architecture and domain adaptation.

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. Currently, model accuracy and generalization ability are the most critical bottlenecks, as evidenced by the performance gaps observed in both multi-domain fusion and cross-domain transfer experiments. To bridge the gap between algorithmic prototypes and reliable deployment, future research must address three pivotal dimensions: robustness, data efficiency, and trustworthy implementation.

A. Robustness and Generalization

The transition from controlled laboratories to unconstrained environments requires systems that are adaptive to both temporal dynamics and domain shifts.

1) Dynamic Temporal Adaptation: Standardizing analysis windows poses a fundamental trade-off; fixed lengths (e.g., 30 seconds) limit the model’s flexibility, as effective detection may require varying temporal resolutions depending on the stressor dynamics. Future frameworks should move towards dynamic window adjustment strategies. For instance, employing Deep Reinforcement Learning agents to autonomously optimize window length can effectively trade off between prediction accuracy and response latency [192].

2) Personalization and Domain Alignment: Inter-subject physiological variability and scenario diversity significantly impede model generalization. Current unsupervised clustering approaches are insufficient for handling these disparities. To build truly effective personalized models, research should leverage transfer learning and fine-tuning strategies, potentially integrating static descriptors (e.g., age, baseline heart rate) to adapt to individual response patterns [193]. Furthermore, our experiments indicate that simple data aggregation may dilute effective signals. Consequently, advanced domain adaptation techniques are essential. Methods such as Maximum Mean Discrepancy (MMD) to align the feature distributions [194] or using contrastive learning and related alignment algorithms (CORAL) to bring different domains closer in the feature space [195] are promising. Recent work combining supervised contrastive learning with metadata demonstrates particular promise in learning generalized stress representations that remain robust to individual differences [196].

B. Data-Centric Paradigms

Overcoming the dual bottlenecks of data scarcity and labeling ambiguity requires a paradigm shift towards generative and self-supervised methodologies.

1) Synthetic Data Augmentation: While we have compiled 21 public datasets, their scale remains insufficient for training complex deep models. Beyond basic physiological modeling [141], the application potential of more advanced generative models, such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models, in this field is yet to be fully explored [197]. Specifically, Conditional Generative Adversarial Networks (cGANs) with diversity constraints have successfully generated high-fidelity EDA signals, preventing mode collapse and providing a blueprint for data augmentation [198].

2) Self-Supervised Learning (SSL): Reliable ground truth remains elusive; activity labels are prone to confounding factors [199], cortisol sampling is invasive [200], and self-reports suffer from subjective bias. SSL offers a pathway to mitigate dependence on labeled data by learning representations from unlabeled signals [201]. While general sensor-based pretext tasks exist, signal-specific strategies—such as learning the dynamic changes of physiological baselines through time-series prediction tasks—prove more effective [202], [203]. Notably, tailoring data augmentation strategies for the EDA signal (e.g., altering frequency components, distorting tonic and phasic components) within contrastive learning frameworks (e.g., SimCLR) has been shown to yield representations that outperform supervised baselines, establishing a new frontier for label-efficient modeling [204].

C. Trustworthy and Efficient Deployment

Bridging the gap between theory and practice necessitates addressing engineering constraints and ethical transparency.

1) Edge Efficiency and Privacy: Deploying models on resource-constrained wearables requires rigorous optimization. Techniques such as TinyML quantization [205] and Neural Architecture Search (NAS)—exemplified by TinyStressNet—can drastically reduce computational overhead without compromising accuracy [142]. Furthermore, bio-inspired Spiking Neural Networks (SNNs) offer a low-power alternative to traditional architectures [206]. Concurrently, data privacy is paramount. Federated Learning (FL) enables decentralized model training, preserving user privacy by keeping raw data local [207], [208], though trade-offs in model performance must be carefully managed [209]. Additionally, ensuring reproducibility across heterogeneous consumer-grade devices remains a critical step for widespread adoption [210].

2) Explainable AI (XAI): To integrate AI into clinical workflows, the "black-box" nature of Deep Learning must be demystified. Post-hoc analysis using Integrated Gradients has confirmed that the rising and recovery phases of the EDA signal's peaks contribute most to stress prediction, aligning with physiological theory [211], while SHAP values validate feature importance in traditional models [212]. Moving forward, Physics-Informed Neural Networks (PINNs) represent a transformative direction. By embedding differential equations governing EDA dynamics into the loss function, PINNs ensure that the internal parameters it learns (such as decay rates) possess clear physiological meaning, thereby fostering the trust required for healthcare applications [213].

VI. Conclusion

This study addresses the systemic fragmentation in stress detection research by presenting the first end-to-end systematic survey of the unimodal EDA pipeline. To foster community reproducibility, we compiled a centralized resource of 21 public datasets and systematically categorized 19 representative unimodal studies, offering a granular analysis of current preprocessing and modeling methodologies. Beyond theoretical review, we established a standardized evaluation protocol and conducted a large-scale empirical benchmark across five diverse datasets. Our results demonstrate that end-to-end DL models significantly outperform traditional feature-engineering approaches, particularly in challenging cross-domain tasks. Crucially, our analysis reveals that the ecological validity of training data is a more decisive factor for generalization than mere data volume, with real-world scenarios yielding more transferable representations than controlled laboratory settings. Building on these insights, we provide a strategic outlook for future advancements. We advocate for a paradigm shift toward advanced techniques such as SSL and Domain Adaptation to overcome persistent challenges in label scarcity and inter-subject variability. Ultimately, by addressing these implementation hurdles, this work guides the field toward the development of more robust, transparent, and deployable health monitoring systems.

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] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[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] P. Sudarsan and A. Niranjil Kumar, "A survey on human stress monitoring technique using electrodermal analysis," *Int. J. Comput. Sci. Eng.*, 2019, pDF available online.

[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, p. 479, 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, p. 7869, 2021.

[38] 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, p. e19, Mar. 2023.

[39] 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. Inform.*, vol. 173, p. 105026, 2023.

[40] J. Kallio, E. Vildjiounaite, J. Tervonen, and M. Bordallo 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.

[41] 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.

[42] 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.

[43] A. Pinge, V. Gad, D. Jaisinghani, S. Ghosh, and S. Sen, "Detection and monitoring of stress using wearables: A systematic review," *Front. Comput. Sci.*, vol. 6, p. 1478851, 2024.

[44] A. Jolly, V. Pandey, M. Sahni, E. Leon-Castro, and L. A. Perez-Arellano, "Modern smart gadgets and wearables for diagnosis and management of stress, wellness, and anxiety: A comprehensive review," *Healthcare*, vol. 13, no. 4, p. 411, 2025.

[45] M. M. Mohod and P. M. Jawandhiya, "A comprehensive literature review of Machine Learning algorithms for stress detection," *Int. J. Sci. Res. Publ.*, vol. 15, no. 2, pp. 88–95, 2025.

[46] A. O. Pataca, E. Zdravevski, P. J. Coelho, N. M. Garcia, M. Deryuck, C. Albuquerque, and I. M. Pires, "Use of Machine Learning for predicting stress episodes based on wearable sensor data: A systematic review," *Comput. Biol. Med.*, vol. 198, p. 111166, 2025.

[47] M. Khandelwal and A. Sharma, "Machine Learning and Deep Learning techniques to detect mental stress using various physiological signals: A critical insight," *WIREs Data Min. Knowl. Discov.*, vol. 15, no. 3, p. e70035, 2025.

[48] A. Abd-Alrazaq, M. Alajlani, R. Ahmad, R. AlSaad, S. Aziz, A. Ahmed, M. Alsahli, R. Damseh, and J. Sheikh, "The performance of wearable AI in detecting stress among students: Systematic review and meta-analysis," *J. Med. Internet Res.*, vol. 26, p. e52622, Jan. 2024.

[49] 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.

[50] 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.

[51] 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.

[52] C. Chatzaki and M. Tsiknakis, "An overview of stress analysis based on physiological signals: Systematic review of open datasets and current trends," *Sensors*, vol. 25, no. 23, 2025. [Online]. Available: https://doi.org/10.3390/s25237108

[53] 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.

[54] 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.

[55] 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.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] 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

[61] 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.

[62] 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

[63] 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.

[64] 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.

[65] 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

[66] 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.

[67] 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.

[68] 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.

[69] 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

[70] 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.

[71] 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.

[72] 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.

[73] 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

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] 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.

[89] 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.

[90] 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.

[91] 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.

[92] 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.

[93] 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.

[94] 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.

[95] 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.

[96] 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.

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

[98] 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.

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

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

[101] 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.

[102] 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.

[103] 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.

[104] 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.

[105] 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.

[106] 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.

[107] 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.

[108] 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.

[109] 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.

[110] 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.

[111] ——, "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.

[112] 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.

[113] 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.

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

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

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

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

[118] 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

[119] 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.

[120] 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.

[121] 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.

[122] 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.

[123] 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.

[124] 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.

[125] 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.

[126] 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.

[127] 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.

[128] 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.

[129] 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.

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

[131] 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.

[132] 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.

[133] S. A. H. Aqajari, E. K. Naeini, M. A. Mehrabadi, S. Labbaf, A. M. Rahmani, and N. Dutt, "Gsr analysis for stress: Development and validation of an open source tool for noisy naturalistic gsr data," 2020. [Online]. Available: https://arxiv.org/abs/2005.01834

[134] 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.

[135] 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.

[136] 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.

[137] 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.

[138] 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.

[139] 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.

[140] 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.

[141] 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.

[142] 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.

[143] 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.

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

[145] 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.

[146] 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.

[147] 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.

[148] 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.

[149] 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.

[150] 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.

[151] 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.

[152] 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.

[153] 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.

[154] 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.

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

[156] 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.

[157] 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.

[158] 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.

[159] 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.

[160] 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.

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

[162] 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.

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

[164] 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.

[165] 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.

[166] 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.

[167] 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

[168] 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.

[169] 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.

[170] 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.

[171] 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.

[172] 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.

[173] H. G. Kim, S. Song, B. H. Cho, and D. P. Jang, "Deep learning-based stress detection for daily life use using single-channel eeg and gsr in a virtual reality interview paradigm," *PLOS ONE*, vol. 19, no. 7, p. e0305864, 2024.

[174] 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.

[175] 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.

[176] 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.

[177] 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

[178] 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.

[179] 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.

[180] 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

[181] 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.

[182] 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.

[183] 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

[184] 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.

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

[186] 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.

[187] 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.

[188] 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.

[189] 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

[190] 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.

[191] 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.

[192] 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.

[193] 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.

[194] 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

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

[196] 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.

[197] 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

[198] 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.

[199] 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.

[200] 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.

[201] 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

[202] 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

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

[204] 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.

[205] M. Mouadili, E. M. En-Naimi, and M. Kouissi, "Advancing stress detection and health monitoring with deep learning approaches," *Computer Sciences & Mathematics Forum*, vol. 10, no. 1, pp. 10–0, 2025. [Online]. Available: https://www.mdpi.com/2813-0324/10/1/10

[206] 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.

[207] 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.

[208] 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

[209] 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.

[210] 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

[211] 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.

[212] 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.

[213] 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