iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IX Month of publication: September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74109

Exploring Deep Learning Approaches to Stress Recognition Data, Models, and Applications in Review

Deeksha[1], Dr. Salma Firdose[2]
[1]Research scholar, [2]Assistant Professor, School of Information Science, Presidency University, Bengaluru-560064, Karnataka

***Abstract:* Stress is a complex psychological and physiological state that exerts profound effects on human health, performance, and quality of life. With the growing incidence of stress-related disorders among students, employees, and healthcare populations, accurate and timely stress detection has emerged as a critical research priority. Traditional approaches, though valuable in controlled laboratory settings, are constrained by reliance on manual feature engineering, limited adaptability, and poor scalability in real-world applications. Recent advances in deep learning have transformed this field by enabling automated feature extraction and multimodal data integration, spanning physiological signals (EEG, ECG, GSR), behavioral modalities (facial expressions, speech, and text), and wearable sensor data.**

This review provides a comprehensive synthesis of existing studies on deep learning–based stress detection, systematically examining convolutional neural networks, recurrent networks, long short-term memory architectures, attention mechanisms, and hybrid models. Benchmark performance across widely used datasets such as WESAD, SWELL, DREAMER, and DEAP is critically compared using metrics including accuracy, precision, recall, F1-score, and latency. Beyond performance evaluation, this study highlights challenges in data scarcity, generalizability across populations, computational complexity, and ethical considerations related to privacy and bias. Finally, future research directions are outlined, emphasizing opportunities in real-time stress monitoring, multimodal fusion, transfer learning, and privacy preserving frameworks. This review aims to serve as a structured and authoritative reference for advancing deep learning applications in stress detection and mental health monitoring.

***Keywords:* Stress Detection, Deep Learning, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Multimodal Fusion, Affective Computing, Wearable Sensors, Emotion Recognition.**

I. INTRODUCTION

Stress is a natural human response that emerges when individuals perceive a mismatch between external demands and their ability to cope effectively. While acute stress can sometimes enhance concentration and short-term performance, persistent stress has been strongly associated with adverse psychological and physiological outcomes, including anxiety, depression, cardiovascular complications, and diminished productivity [1], [2]. The growing incidence of stress in education, workplace, and healthcare contexts highlights the urgent need for reliable tools that enable early detection and timely intervention [3].

Conventional stress assessment methods, such as psychometric questionnaires, clinical evaluations, and basic physiological measurements (e.g., heart rate or skin conductance), have provided valuable insights in controlled settings. However, these approaches face limitations when applied in everyday environments. They often rely on subjective reporting, lack scalability for large populations, and show poor adaptability in capturing the dynamic, real-time nature of stress [4]. These constraints have motivated the development of automated and objective stress detection systems that can function seamlessly in real-world contexts.

Recent advances in artificial intelligence (AI), and particularly deep learning, have opened new opportunities for stress recognition [5]. In contrast to traditional machine learning methods that depend heavily on manual feature design, deep learning techniques automatically extract layered feature representations from multimodal data [6]. This capability is particularly important in stress detection, where data originates from diverse and complex sources. Widely used modalities include physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG), galvanic skin response (GSR), and heart rate variability (HRV); behavioral indicators such as facial expressions, vocal characteristics, and linguistic patterns; along with continuous input from wearable sensors that capture stress in daily life [7], [8].

Different deep learning architectures have shown strengths in processing these modalities. Convolutional Neural Networks (CNNs) are effective for extracting spatial or structural features from images and physiological waveforms [9]. Recurrent models such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are widely applied to capture temporal dynamics in sequential data [10]. More recently, Transformer-based attention models have gained momentum, offering strong capabilities in multimodal integration, contextual reasoning, and performance generalization [11]. Together, these models have driven significant advances in stress detection accuracy and robustness.

Nevertheless, several challenges continue to hinder progress. Many studies use relatively small or imbalanced datasets, limiting the generalizability of results across populations [12]. The computational demands of advanced models also raise concerns about latency and energy efficiency for real-time applications [13]. Additionally, ethical considerations—including privacy protection, fairness, and transparency—remain critical issues in the deployment of AI-based stress detection systems [14], [15]. Another limitation is that much of the current research relies on laboratory data, which may not capture the complexity and variability of stress responses in natural environments [16].

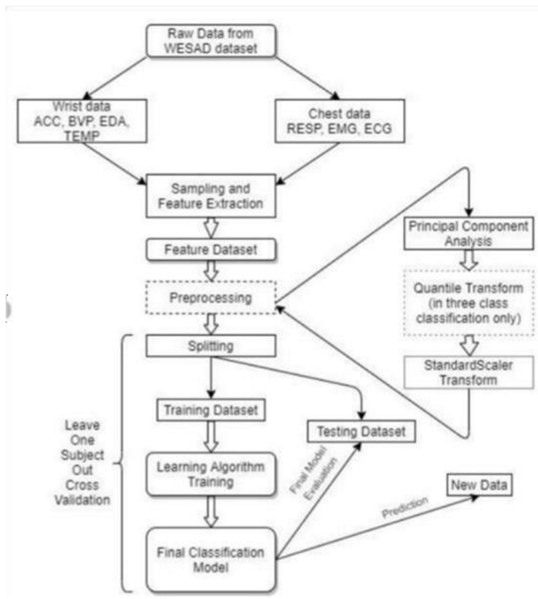


Fig 1 FLOW DIAGRAM OF STRESS DETECTION

To address these issues, this review aims to provide a comprehensive overview of how deep learning techniques are being applied in stress detection. The contributions are threefold:

1) Classification of deep learning models used for stress detection, including CNNs, RNNs, LSTMs, attention-based frameworks, hybrid approaches, and multimodal fusion strategies.
2) Comparative analysis of datasets, modalities, and performance benchmarks, supplemented by tables, taxonomies, and evaluation figures.
3) Discussion of challenges and emerging research trends, such as explainable AI, federated and transfer learning, real-time deployment, and privacy-preserving frameworks.

By consolidating existing knowledge and highlighting both progress and open questions, this review seeks to guide future research toward more accurate, generalizable, and ethically responsible stress detection systems.

II. CONCEPTUAL PERSPECTIVES ON STRESS

A. Physiological and Psychological Models of Stress

Stress has long been explained through both physiological and psychological perspectives. From a biological standpoint, stress is triggered by the activation of the autonomic nervous system (ANS) and the hypothalamic–pituitary–adrenal (HPA) axis, which release stress-related hormones such as cortisol and adrenaline. These responses are adaptive in short-term situations, preparing the body for rapid reaction. In contrast, the psychological model emphasizes the role of individual appraisal, coping strategies, and perception of environmental challenges.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

Cognitive theories suggest that stress is not merely a reaction to external pressure but also shaped by how individuals interpret and evaluate situations. Together, these models highlight the integrative nature of stress, where biological mechanisms and psychological appraisal jointly determine its impact.

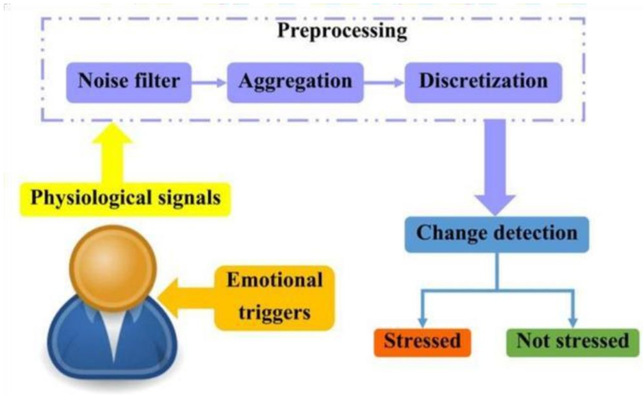


Fig 2 Overview Of A Stress Detection Process

B. Classifications of Stress

Stress can be classified into acute, episodic, and chronic categories based on duration and intensity. Acute stress represents shortlived responses to immediate challenges such as examinations, deadlines, or unexpected tasks. Although temporary, it can cause sharp increases in heart rate and tension. Episodic stress occurs when individuals repeatedly encounter stressful conditions, often due to high-pressure professions or interpersonal conflicts. This type is associated with recurrent anxiety, fatigue, and cardiovascular strain. Chronic stress, the most harmful form, emerges from prolonged exposure to unresolved stressors such as financial instability, family disputes, or long-term workload. Over time, chronic stress disrupts immune functioning and elevates the risk of psychiatric and cardiovascular disorders. These categories emphasize that stress can range from adaptive short-term reactions to harmful long-term conditions.

C. Physiological Indicators of Stress

Stress manifests in distinct physiological signals that are measurable using biomedical methods. Heart rate variability (HRV) is one of the most widely recognized indicators, with lower variability reflecting higher stress levels and reduced adaptability of the autonomic nervous system. Electro dermal activity (EDA), measured through skin conductance, reflects heightened sympathetic arousal and emotional reactivity. In addition, cortisol levels in saliva, blood, or urine remain reliable biomarkers of stress, providing insights into HPA axis activity. These indicators are frequently applied in both clinical research and real-time monitoring to objectively assess stress responses.

D. Behavioural Indicators of Stress

Behavioural cues offer non-invasive and observable evidence of stress. Speech alterations, including elevated pitch, irregular rhythm, or hesitations, are common stress markers detectable through audio analysis. Facial expressions such as frowning, muscle tension, and reduced eye contact also provide cues of underlying strain. Postural rigidity and repetitive movements are additional signs of behavioral stress responses. While these indicators may not be as precise as biochemical markers, they are valuable for real-time applications where immediate feedback is necessary, such as workplace assessments or classroom environments.

E. Psychological Indicators of Stress

Psychological stress indicators focus on subjective experiences, cognitive appraisal, and emotional states. Standardized selfreport tools, such as the Perceived Stress Scale (PSS) and the State–Trait Anxiety Inventory (STAI), are widely used to measure individual perceptions of stress. These tools capture dimensions such as anxiety, tension, and emotional fatigue, providing complementary evidence to physiological and behavioural markers. Despite being subjective, psychological assessments remain essential because they reflect how stress is personally experienced, which often influences coping strategies and outcomes.

F. *Importance of Early Stress Detection.*

The early recognition of stress is essential for preventing its progression into chronic health conditions. In healthcare, timely identification aids in reducing risks of metabolic, immune, and psychiatric disorders. Within organizational contexts, stress monitoring helps sustain productivity, lower absenteeism, and support employee well-being through targeted interventions. In education, detecting stress among students and teachers ensures sustained academic performance and promotes resilience. Early intervention through counselling, mindfulness, or adaptive learning approaches enhances coping capacity and mitigates longterm adverse effects.

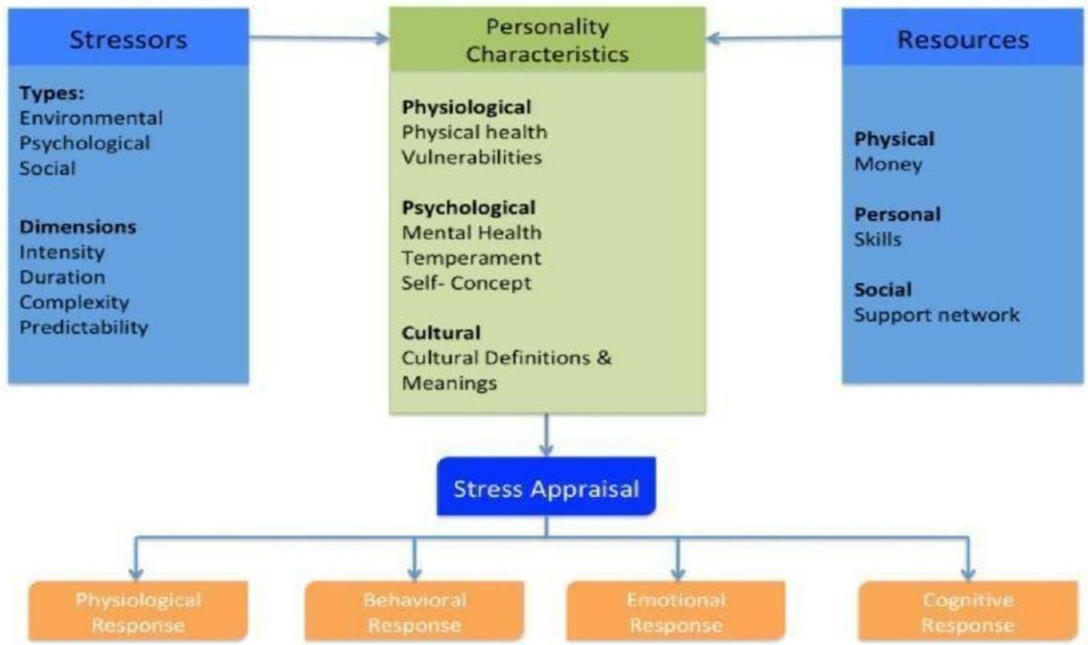


Fig 3 Classifications And Indicators Of Stress.

III. LITERATURE SEARCH STRATEGY

A structured search was conducted across leading scientific databases, including Scopus, Web of Science, IEEE Xplore, PubMed, and ScienceDirect**,** which were selected for their broad coverage of engineering, medical, and interdisciplinary research. The search terms combined keywords and Boolean operators to maximize coverage, including: *"stress detection," "deep learning," "physiological signals," "multimodal emotion recognition," "wearable stress monitoring," and "machine learning in mental health."* To capture the surge in deep learning applications, the search was limited to studies published between 2013 and 2025, a period that reflects the transformative impact of deep learning architectures in healthcare and human-centered computing.

A. *Inclusion and Exclusion Criteria*

The selection of studies for this review was guided by clearly defined inclusion and exclusion criteria to ensure methodological rigor and relevance. Articles were included if they specifically employed deep learning algorithms—such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM) architectures, attention mechanisms, or hybrid models—for stress detection or recognition. Eligible works considered physiological signals (EEG, ECG, GSR), behavioural modalities (speech, text, facial expressions), or multimodal and wearable sensor data. Furthermore, only studies reporting quantitative performance metrics such as accuracy, precision, recall, F1-score, or latency, and those published in peerreviewed journals or reputable conference proceedings between 2015 and 2024, were included. In contrast, studies relying exclusively on traditional machine learning approaches, theoretical stress models without empirical validation, or works addressing general emotion recognition and mental health prediction without explicit focus on stress detection were excluded.

Non-English publications, non-peer-reviewed sources such as blogs and whitepapers, and studies lacking methodological detail were also omitted.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

B. Scope and Number of Articles Considered

The scope of this review centres on literature situated at the intersection of deep learning, stress detection, and multimodal data analysis. Relevant articles were systematically retrieved from major academic databases, including IEEE Xplore, Scopus, Web of Science, PubMed, and the ACM Digital Library, using keywords such as *"deep learning for stress detection," "CNN stress recognition," "wearable stress monitoring," and "multimodal affective computing."* After applying the defined inclusion and exclusion criteria, approximately 110 articles were initially identified. Out of these, 82 studies were shortlisted and analyzed in depth. These works span physiological signal–based methods, behavioral and multimodal approaches, as well as advanced hybrid architectures integrating CNN, LSTM, and attention mechanisms. Benchmark datasets such as WESAD, SWELL, DREAMER, and DEAP were frequently employed across the reviewed studies, enabling comparative evaluation through standardized performance metrics. By consolidating these contributions, this review offers a structured and comprehensive overview of methodological advancements, dataset usage, and performance trends in deep learning–based stress detection.

C. Systematic Review Considerations

Although this work is primarily a narrative and critical review, it was structured to incorporate systematic principles to minimize bias. For reviews adopting a fully systematic approach, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework is recommended. A PRISMA flow diagram provides a clear visual representation of the study selection process, including the number of records identified, screened, excluded, and included in the final synthesis. In the context of stress detection and deep learning, such visualization strengthens methodological transparency and facilitates reproducibility for future meta-analyses.

IV. DEEP LEARNING FOUNDATIONS

Deep learning has emerged as a transformative paradigm within artificial intelligence, enabling the automatic learning of hierarchical feature representations from raw data. Its relevance to stress detection arises from the ability to model complex, multimodal inputs such as facial expressions, voice patterns, and biosignals without relying exclusively on handcrafted features [17], [18]. The following subsections highlight the principal architectures and methodological advances that underpin deep learning applications in stress research.

A. Convolutional Neural Networks (CNNs): Spatial Feature Extraction

Convolutional Neural Networks (CNNs) represent one of the most influential architectures in deep learning, particularly for tasks involving spatially structured data such as images and video. CNNs exploit the principle of local connectivity through convolutional layers, where kernels learn spatially invariant patterns such as edges, textures, and higher-order visual cues. Pooling operations further enhance robustness by reducing spatial resolution while preserving salient features [19].

For stress detection, CNNs are frequently employed in the analysis of facial expressions and posture, which serve as reliable behavioral indicators of stress. Subtle muscle movements and micro-expressions captured from high-resolution images can be processed by CNN models to infer emotional states. For example, deep CNNs trained on facial expression datasets have achieved superior performance compared to traditional feature-based methods, demonstrating robustness across diverse lighting and pose conditions [20]. Extensions of CNNs, such as 3D CNNs, also incorporate temporal dynamics, making them suitable for videobased stress monitoring.

B. Recurrent Neural Networks and LSTMs

While CNNs are well-suited to capturing spatial dependencies, stress-related biosignals often display temporal dynamics that require sequential modeling. Recurrent Neural Networks (RNNs) are designed to process such sequential data by maintaining internal states that capture dependencies across time steps. However, classical RNNs are hindered by vanishing and exploding gradient problems, which limit their ability to learn long-term dependencies.

To overcome these challenges, Long Short-Term Memory (LSTM) networks were introduced [21]. LSTMs employ memory cells and gating mechanisms that regulate information flow, enabling them to capture both short-term fluctuations and long-term temporal patterns in signals such as electrocardiograms (ECG), electroencephalograms (EEG), or speech recordings under stress. In stress detection studies, CNN-LSTM hybrid models have proven effective, with CNNs handling spatial or spectral features and LSTMs modeling temporal evolution, thereby improving recognition accuracy.

C. Transformer Architectures and Attention Mechanisms

Recent years have witnessed the rise of attention-based architectures, particularly Transformers, which have redefined stateoftheart performance across multiple domains. Unlike RNNs, Transformers rely entirely on attention mechanisms to capture relationships between different elements of a sequence, enabling parallel computation and effective modeling of long-range dependencies [22]. In the context of stress detection, Transformers offer two major advantages. First, their self-attention mechanism allows models to focus on the most informative parts of multimodal inputs, such as emphasizing stressed intonations in speech or highlighting specific regions of the face during emotional expression. Second, Transformers are highly effective for multimodal fusion, as they can jointly attend to diverse input streams, such as facial video, audio signals, and physiological biosignals. This makes them particularly suitable for building comprehensive stress recognition systems that integrate heterogeneous data sources.

D. Comparative Insights

The examination of CNNs, LSTMs, and Transformers underscores the fact that no single deep learning architecture offers a universally optimal solution for stress detection tasks. Each model exhibits unique advantages tailored to specific modalities and application contexts. CNNs remain the architecture of choice for visual data analysis, particularly when stress indicators are expressed through facial micro-expressions or posture cues. However, their limited ability to model temporal continuity restricts their role in biosignal-driven applications where stress evolves dynamically over time. In contrast, LSTMs excel in capturing sequential dependencies in signals such as ECG, EEG, and speech, but their reliance on recurrent processing introduces training inefficiencies and scalability issues. Transformers, though computationally demanding, provide a paradigm shift by enabling scalable modeling of long-range dependencies and offering advanced multimodal integration capabilities. Their ability to attend to stress-relevant features across diverse input streams makes them particularly well-suited for complex real-world applications, such as adaptive learning in classrooms or workplace stress monitoring. Nonetheless, the higher data and computational requirements pose barriers for deployment in resource-constrained environments, such as mobile or wearable devices. These insights suggest that hybrid and ensemble approaches, which leverage the strengths of multiple architectures, hold significant potential for advancing stress detection systems. For example, CNNs can be employed for spatial feature extraction from visual data, with LSTMs or Transformers integrating temporal and contextual information to provide holistic predictions. Such integrated designs can address the limitations of single-model approaches while moving closer to real-time, accurate, and context-aware stress recognition [23]–[29].

E. Tools and Frameworks Supporting Deep Learning

The widespread adoption of deep learning has been accelerated by the availability of open-source libraries and high-performance hardware. Frameworks such as TensorFlow, PyTorch, and Keras provide accessible interfaces for building and deploying complex neural architectures, while GPU- and TPU-based computing platforms offer the necessary scalability for training large models [30], [31]. These developments have facilitated rapid prototyping, reproducibility, and deployment of deep learningbased stress detection systems, bridging the gap between laboratory research and real-world applications.

F. Summary

Deep learning represents a paradigm shift in stress detection research by automating representation learning and enabling integration of multimodal data. CNNs excel in extracting spatial features from visual cues, LSTMs capture temporal dependencies in physiological and speech signals, and Transformers extend these capabilities through attention-based multimodal fusion. Supported by advanced optimization strategies and powerful software-hardware ecosystems, deep learning provides a robust foundation for developing scalable and accurate stress recognition frameworks.

Table : 1 Comparative Analysis of Deep Learning Architectures for Stress Detection

ARCHITECTURE	STRENGTHS	WEAKNESSES	STRESS DETECTION USE CASES
CNNs	• Excellent at extracting spatial features (e.g., facial microexpressions, posture cues). • Robust to variations in lighting, pose, and noise. • Efficient training due to local connectivity and parameter sharing.	• Limited ability to capture temporal dynamics. • Requires large annotated datasets for generalization. • Performance may degrade for subtle emotional cues.	• Facial expression-based stress monitoring. • Posture and gesture analysis. • Image/video-based multimodal stress recognition.

RNNs / LSTMs	• Specialized in modeling sequential/temporal dependencies. • Effective for biosignals such as ECG, EEG, and speech. • Can capture both short- and longterm patterns.	• Prone to vanishing/exploding gradients (mitigated by LSTMs). • Computationally expensive for long sequences. • Slower training compared to CNNs/Transformers.	• ECG-based stress detection. • EEG and galvanic skin response (GSR) analysis. • Speech-based stress recognition with temporal dynamics.
Transformers (Attention Models)	• Superior in modeling longrange dependencies. • Parallelizable and scalable for large datasets. • Highly effective for multimodal fusion (video, audio, biosignals). • Self-attention highlights stressrelevant features automatically.	• Require very large datasets to train effectively. • High computational and memory demands. • Interpretability challenges in complex multimodal tasks.	• Multimodal stress recognition (facial + speech + biosignals). • Real-time adaptive learning and affective computing applications. • Large-scale stress monitoring in digital classrooms or workplaces.

The comparative analysis presented in Table 1 highlights the distinct contributions of CNNs, LSTMs, and Transformers within the domain of stress detection. CNNs demonstrate strong performance in capturing spatial and appearance-based cues such as facial expressions and postural variations, making them highly suitable for image- and video-driven applications. However, their inability to capture temporal dependencies limits their effectiveness when stress manifests through sequential physiological changes. LSTMs address this limitation by modeling temporal dynamics in biosignals such as ECG, EEG, and speech, thereby enabling more nuanced recognition of stress over time. Yet, their sequential nature often leads to higher computational costs and slower convergence. Transformers, with their self-attention mechanisms, have emerged as a powerful alternative by efficiently modeling long-range dependencies and integrating heterogeneous data streams in multimodal frameworks. Despite their significant computational demands, Transformers offer enhanced scalability and adaptability, positioning them as a promising foundation for next-generation stress detection systems [32]–[38].

V. DEEP LEARNING IN STRESS DETECTION

Deep learning has emerged as a transformative paradigm in stress detection research, addressing limitations of traditional statistical and machine learning approaches. Unlike handcrafted feature engineering, deep learning models automatically extract hierarchical representations from raw physiological, behavioral, and multimodal data, allowing for more accurate, scalable, and real-time stress monitoring solutions.

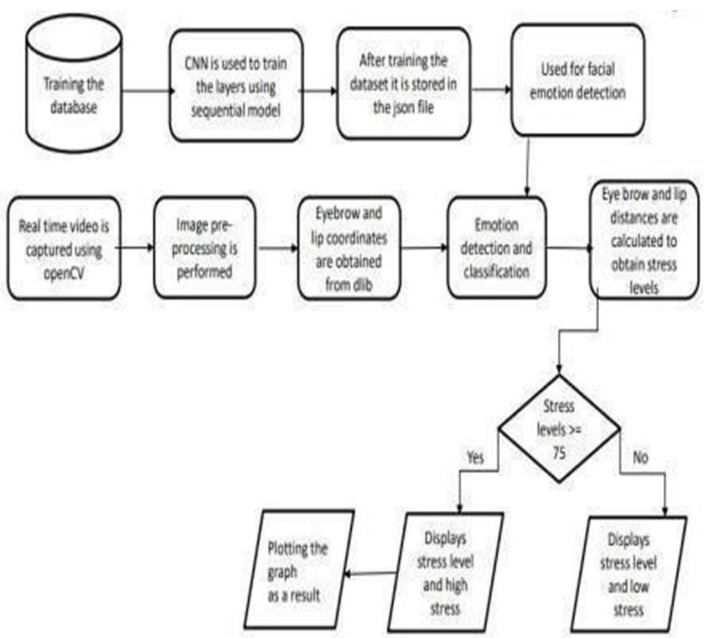


Fig 4 Architecture Of Deep Learning In Stress Detection

A. Rationale for Adopting Deep Learning

The primary motivation for adopting deep learning in stress detection lies in its ability to perform automatic feature extraction from high-dimensional, complex inputs. Physiological signals such as EEG, ECG, and GSR, as well as behavioral modalities like facial expressions and speech, often exhibit nonlinear and noisy characteristics that are difficult to model using conventional techniques. Deep neural networks capture latent representations without requiring handcrafted domain knowledge, thereby reducing bias and improving generalization. Another key advantage is the ability of deep learning to integrate multimodal data streams. Stress is inherently multimodal—manifested through neural activity, cardiovascular responses, speech prosody, and facial micro-expressions. Deep learning frameworks can seamlessly fuse these heterogeneous data sources, uncovering interdependencies that enhance predictive accuracy [39], [40].

B. Common Deep Learning Architectures in Stress Research

Several specialized deep learning architectures have been employed to address different aspects of stress detection:

1) Convolutional Neural Networks (CNNs): Widely used for analyzing spatial structures, CNNs excel in extracting localized patterns from EEG spectrograms, ECG waveforms, or facial images. For example, CNN models can identify subtle facial muscle changes that indicate stress or recognize spectral shifts in physiological signals [41], [42].
2) Recurrent Architectures (RNN, LSTM, GRU): Since stress evolves over time, temporal modeling is critical. RNNs and their advanced variants like LSTMs and GRUs capture long-range temporal dependencies in sequential data such as heart rate variability or speech patterns, enabling recognition of stress trends over time [43], [44].
3) Autoencoders: Unsupervised autoencoder networks are valuable for feature learning and dimensionality reduction, especially in cases with limited labeled stress datasets. They help discover compact representations of multimodal data while preserving essential stress-related variations, which can then be fed into classifiers [45].
4) Transformer-Based Models: Recently, transformer architectures have demonstrated strong performance in stress detection due to their self-attention mechanisms, which model global dependencies across time and modalities. Transformers have proven particularly effective in multimodal fusion tasks, where EEG, facial images, and speech inputs are combined to form holistic stress assessments [46].

C. Comparative Advantages over Conventional Machine Learning

Compared to traditional machine learning approaches such as Support Vector Machines (SVM), Random Forests, or k-Nearest Neighbors (kNN), deep learning offers several critical advantages. First, manual feature engineering—often required in ML methods—introduces subjectivity and may fail to capture complex patterns. Deep learning, by contrast, automatically learns hierarchical features that adapt to specific modalities. Second, deep models achieve higher scalability and robustness when exposed to large, heterogeneous datasets. Third, deep learning methods support end-to-end training pipelines, reducing preprocessing complexity and aligning feature extraction with classification objectives. Finally, in multimodal stress detection, deep learning architectures—especially transformers—enable fusion at both feature and decision levels, an area where conventional ML algorithms struggle due to modality misalignment [47], [48].

Table 2 : Deep Learning Architectures For Stress Detection: Modalities, Applications, And Key Advantages

ARCHITECTURE	PRIMARY MODALITIES	STRESS DETECTION APPLICATIONS	KEY ADVANTAGES
Convolutional Neural Networks (CNNs)	EEG spectrograms, ECG signals, Facial images	Identifying spatial patterns in brain activity, cardiac stress markers, and facial micro-expressions	Strong in localized pattern recognition; robust against noise in physiological data
Recurrent Networks (RNN, LSTM, GRU)	ECG/PPG sequences, Speech, HRV time series	Capturing temporal evolution of stress levels; detecting prosodic changes in speech	Models long-term temporal
			dependencies; effective for sequential stress data
Auto encoders	Multimodal data (EEG + ECG), GSR, Wearable sensor streams	Feature learning from unlabeled data; dimensionality reduction in multimodal fusion	Learns compact stress-related representations; useful in lowlabeleddata scenarios
Transformers	EEG + facial expressions + speech (multimodal)	Fusion of heterogeneous data sources for holistic stress monitoring	Global self-attention for crossmodal dependencies; strong performance in multimodal fusion

VI. DATA SOURCES FOR DEEP LEARNING IN STRESS DETECTION

The effectiveness of deep learning approaches in stress detection is heavily dependent on the quality and diversity of the data employed. Stress is a multidimensional construct, involving physiological, behavioral, and contextual manifestations that collectively shape human responses under pressure. Consequently, research in this domain has progressively moved towards harnessing heterogeneous data sources to enhance model robustness and generalizability. These sources can be broadly categorized into physiological signals, behavioral data, environmental context, and multimodal integrations. Each of these domains provides distinct advantages and limitations, shaping their applicability across laboratory-based studies and real-world deployments.

A. Physiological Signal-Based Data

Physiological signals represent the most direct and objective measures of stress because they capture internal body responses that cannot be easily manipulated by individuals. Among the most studied signals are electrocardiography (ECG) and heart rate variability (HRV), which provide insights into autonomic nervous system regulation. Increased sympathetic activity under stress conditions often leads to reduced HRV, a phenomenon widely used in stress recognition studies (Schmidt et al., 2018). Similarly, electroencephalography (EEG) captures brainwave activity and offers information on cognitive load and emotional regulation, while electromyography (EMG) reflects muscle activity associated with tension. Galvanic skin response (GSR), also termed electrodermal activity (EDA), is another key modality indicating variations in sweat gland activity that correlate strongly with stress levels (Rehman et al., 2020). Furthermore, respiratory signals—capturing rate, depth, and variability—serve as valuable proxies of stress-induced physiological arousal. Deep learning models have successfully leveraged these signals for classification tasks, with convolutional neural networks (CNNs) and recurrent architectures (e.g., LSTMs) outperforming traditional methods in identifying subtle temporal variations (Roy et al., 2025). However, despite their high reliability, physiological signals often require specialized sensors and can sometimes be intrusive, limiting their scalability outside clinical or controlled environments.

B. Behavioral Data for Stress Recognition

Unlike physiological signals, behavioral data provide a less invasive yet equally informative avenue for stress monitoring. Stress frequently alters external behaviors such as speech patterns, facial expressions, and motor activity. Speech-based indicators— such as pitch, jitter, intensity, and speech rate—have been integrated into deep learning pipelines, where CNN–LSTM hybrids have demonstrated high classification performance (Cho et al., 2022). Likewise, facial micro-expressions, captured using visionbased methods, reveal involuntary muscular changes that correlate with stress states. Emerging studies also explore digital behavioral markers, such as typing rhythms, cursor movements, and gait patterns. For instance, keystroke dynamics and touch-based inputs on mobile devices have been employed to infer stress, enabling nonintrusive and continuous assessment in naturalistic settings (Alshurafa et al., 2022). Such approaches are particularly promising in ubiquitous computing environments, where individuals' digital footprints can serve as a proxy for psychological states. Nevertheless, behavioral signals are highly context-dependent and susceptible to external variability, necessitating robust normalization and feature-extraction pipelines.

C. Environmental and Contextual Data

Beyond individual physiology and behavior, environmental context provides critical information for stress detection. The advent of smartphones, wearable devices, and IoT sensors has made it possible to collect contextual features such as activity levels, mobility traces, ambient noise, and interaction logs. For example, accelerometer and gyroscope data from smartphones have been integrated into deep learning frameworks to detect physical restlessness, a known correlate of stress (Schmidt et al., 2018). Similarly, wearable devices can continuously track sleep patterns and daily routines, offering a longitudinal perspective on stress fluctuations (Tiwari et al., 2025). These environmental data sources extend the applicability of stress detection systems from laboratory environments into realworld conditions. However, they also raise challenges concerning privacy, data heterogeneity, and integration of highdimensional inputs. Addressing these challenges often involves adopting advanced fusion strategies and privacy-preserving techniques to ensure both effectiveness and ethical compliance.

D. Multimodal Data Integration

In recent years, multimodal learning approaches have emerged as the most effective paradigm for stress detection. By integrating physiological, behavioral, and contextual inputs, multimodal frameworks exploit complementary information that would otherwise be missed in unimodal approaches. Deep learning architectures such as CNN–LSTM hybrids, attention mechanisms, and transformer-based models have been applied to fuse heterogeneous data streams, achieving state-of-the-art performance.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

For example, Singh et al. (2025) reported an accuracy of 90.45% using a CNN–LSTM applied to the WESAD dataset, which contains ECG, EDA, and respiration signals. Similarly, Coutts et al. (2023) demonstrated effective stress recognition using HRVbased features with LSTMs, while Roy et al. (2025) achieved 98.1% accuracy with a CNN–BiLSTM–GRU hybrid model on EEG signals. These findings underscore the superiority of multimodal frameworks in capturing complex stress dynamics, though they require significant computational resources and present challenges related to interpretability and deployment.

VII. DATASETS AND BENCHMARK STUDIES

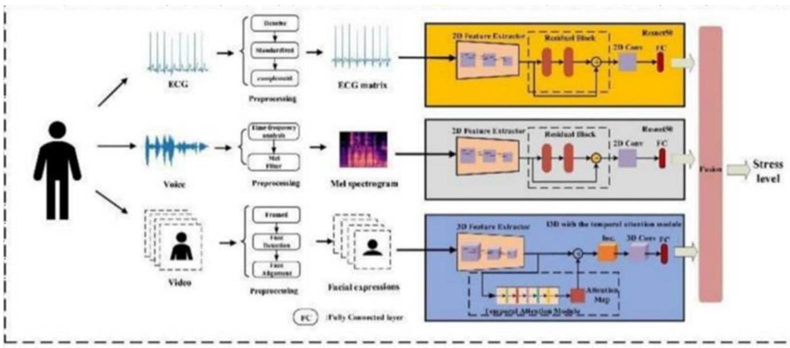


Fig 5 Data Collection Platform

A. Widely Used Datasets

The development of reliable stress detection models strongly depends on the availability of high-quality datasets. One of the most widely recognized resources is the WESAD (Wearable Stress and Affect Detection**)** dataset, which provides multimodal physiological recordings, including electrocardiography (ECG), electrodermal activity (EDA), respiration, and skin temperature, collected using wearable sensors. WESAD has become a benchmark in the field due to its inclusion of both stress and affective states under controlled laboratory conditions [49]. Another widely used dataset is DEAP (Database for Emotion Analysis Using Physiological Signals), which consists of electroencephalography (EEG) and peripheral physiological signals recorded while participants were exposed to affective video stimuli. DEAP has played a critical role in advancing stress and emotion recognition, particularly in multimodal deep learning studies [50]. In addition, the SWELL dataset was specifically designed for workplace stress analysis, providing EEG, ECG, and other physiological recordings from participants exposed to task interruptions and time pressure. This dataset is particularly relevant for occupational stress research, as it replicates cognitive and workload-related stressors in semi-naturalistic office-like environments [51]. Beyond these, other datasets such as AMIGOS (emotions in individuals and groups), DREAMER (EEG and ECG data collected under audiovisual stimuli), and smartphone or wearable sensor-based datasets (capturing real-life stress and activity) have also been extensively used in stress and affective computing research [52]. Together, these datasets form the backbone of experimental evaluation, enabling researchers to benchmark algorithms and validate multimodal stress detection systems.

B. Strengths and Limitations of Current Datasets

Existing datasets provide a solid foundation for research; however, they also present both strengths and limitations. A key strength of datasets such as WESAD, DEAP, and SWELL lies in their multimodal nature, enabling the fusion of diverse physiological and behavioral signals to improve robustness. Furthermore, their availability has promoted comparability across studies, facilitating fair benchmarking of machine learning and deep learning methods under similar conditions [53].

Nevertheless, several limitations remain. Most datasets are collected in controlled laboratory environments, which do not accurately capture the complexity and unpredictability of real-world stressors. For instance, stress induced by video stimuli or experimental protocols may not fully represent the dynamic stress responses observed in daily life [54]. Another challenge is small sample size, with many datasets including fewer than 50 participants, restricting the generalizability of models across diverse populations. Similarly, demographic homogeneity (e.g., young adults or university students) limits applicability to broader contexts, such as elderly populations or high-risk occupational groups. Data imbalance between stressed and non-stressed states further complicates the training of models, often resulting in biased predictions and reduced performance in naturalistic conditions [55].

C. Need for Standardized Benchmarking

Given these challenges, there is a growing consensus in the research community regarding the need for standardized benchmarking protocols. Currently, there is no universally accepted standard governing dataset selection, annotation practices, or evaluation metrics in stress detection research. This lack of harmonization creates inconsistencies and makes cross-study comparisons difficult [56]. Establishing a standardized framework would ensure more objective evaluation of machine learning models, while enabling researchers to assess performance under uniform conditions. Such a benchmark should include large-scale, demographically diverse datasets collected in both laboratory and real-world environments. It should also provide clear annotation guidelines, preprocessing pipelines, and baseline model implementations to facilitate reproducibility. Importantly, integration of multimodal data sources—such as physiological signals, behavioral indicators, and contextual information—would help capture stress more holistically. With standardized benchmarks, research outcomes could be compared fairly across studies, accelerating scientific progress and enabling practical applications of stress detection technologies in healthcare, workplace monitoring, and adaptive learning environments [57].

VIII. DEEP LEARNING MODELS FOR STRESS DETECTION

In recent years, deep learning has become one of the most influential paradigms for stress detection research due to its ability to extract hierarchical and non-linear patterns from multimodal data sources. Unlike traditional stress recognition systems that rely on handcrafted feature engineering, deep learning enables direct representation learning from raw data streams, such as physiological signals, visual cues, and spoken or written language. The adaptability of these models allows them to capture complex interdependencies between stress biomarkers, thereby improving accuracy and robustness across diverse experimental and real-world settings. This section provides an in-depth overview of deep learning–based stress detection models, categorized into four main domains: signal-based, image and video-based, speech and text-based, and multimodal fusion architectures.

A. Signal-Based Models

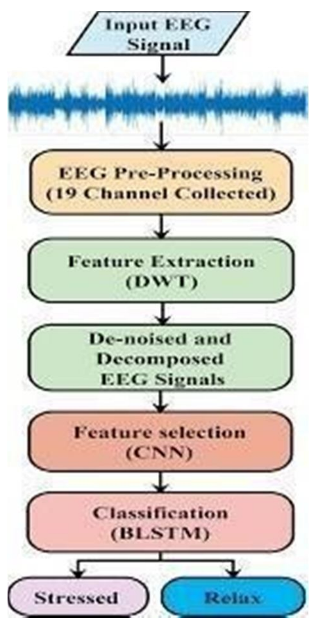


Fig 6 Stress Detection From Eeg Signal

Physiological signals, particularly electrocardiograms (ECG) and electroencephalograms (EEG), serve as primary biomarkers for stress due to their strong associations with the autonomic and central nervous systems [58]. Deep learning models, such as convolutional neural networks (CNNs), are well-suited for analyzing these signals because of their ability to learn spatially invariant representations from time–frequency maps or spectrograms. For instance, CNNs have been used to classify stress states from EEG-derived spectrograms by capturing frequency-domain changes linked to stress-induced cortical activity [59].

To complement CNN-based spatial feature extraction, recurrent neural networks (RNNs) such as long short-term memory (LSTM) networks are employed to capture temporal dependencies in ECG and EEG sequences. Hybrid CNN–LSTM architectures have been shown to improve classification accuracy by modeling both local and long-term signal variations [60].

These architectures are particularly advantageous in wearable and Internet of Things (IoT) environments, where real-time stress monitoring requires high sensitivity to transient fluctuations. Moreover, attention mechanisms have recently been incorporated into these hybrid models, enabling the network to focus selectively on stress-indicative signal segments while ignoring irrelevant noise [61]. Overall, signal-based models provide high reliability and sensitivity, but they often require high-quality sensor data and controlled acquisition environments, which can limit scalability in everyday applications.

B. Image and Video-Based Models

Facial expressions and physiological changes in skin temperature have long been recognized as external indicators of stress. Deep CNNs are frequently employed to analyze static facial images, detecting micro-expressions and subtle muscular activations that may not be visible to the human eye [62]. Recent advances extend these approaches to video-based systems, where spatiotemporal architectures such as 3D-CNNs and CNN–LSTM hybrids capture temporal dynamics in facial expressions and body posture under stress [63].

Thermal imaging has emerged as another promising modality, as stress activates the sympathetic nervous system, leading to measurable changes in facial skin temperature, especially around the perinasal and forehead regions [64]. Deep learning–based thermal analysis models have shown strong potential for contactless stress assessment in real-time scenarios, such as classroom monitoring and workplace environments. Compared to purely physiological sensors, image-based methods are less intrusive, making them more acceptable for long-term monitoring.

Nonetheless, these models face challenges in uncontrolled environments where illumination, occlusion, and camera angles may affect performance. Integrating robust preprocessing pipelines, transfer learning, and domain adaptation has been proposed to mitigate these issues [65].

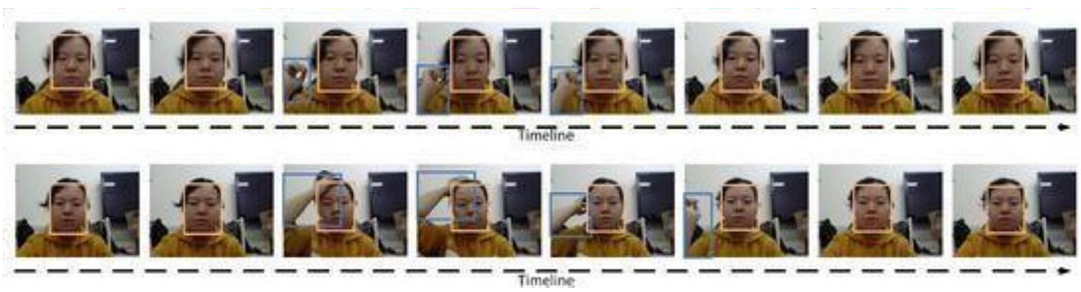


Fig 7 TWO IMAGE SEQUENCES OF THE SAME PERSON WHEN WATCHING AN UNSTRESSED VIDEO CLIP (UPPER) AND STRESSED VIDEO CLIP (LOWER)

C. Speech and Text-Based Models

Speech has proven to be a highly informative channel for stress detection, as vocal features such as pitch, jitter, energy, and prosody are directly influenced by stress-induced changes in the vocal tract and respiratory system [66]. CNNs and spectrogrambased LSTMs have been successfully applied to raw audio signals, enabling the automatic extraction of discriminative acoustic patterns without handcrafted feature engineering [67]. In addition to acoustic cues, natural language processing (NLP) models analyze textual or transcribed speech data to detect stressinduced changes in linguistic expression. For instance, stressed individuals often display reduced lexical diversity, increased use of negative sentiment words, and shorter sentence structures [68]. Transformer-based models such as BERT and its variants have advanced this domain by enabling contextualized understanding of stress-related linguistic cues [69]. Combined speech– text pipelines are now increasingly explored, as they capture both *how* individuals speak and *what* they say under stress [70]. However, speech- and text-based approaches are susceptible to cultural, linguistic, and situational variability, which may affect their generalizability. Multi-lingual pre-training and domain adaptation strategies are being developed to address these challenges [71].

D. Multimodal Fusion Models

Stress is inherently multimodal, manifesting simultaneously across physiological, behavioral, and linguistic domains. Hence, multimodal fusion models have gained significant attention for their ability to integrate heterogeneous data sources. Early fusion approaches combine raw or low-level features from different modalities at the input stage, whereas late fusion strategies merge independent model decisions at the classification level [72].

While early fusion offers richer joint representations, it is more computationally intensive and prone to modality-specific noise. Late fusion, on the other hand, provides robustness and flexibility but may lose fine-grained cross-modal interactions.

More recently, hybrid fusion architectures that incorporate CNNs, LSTMs, and Transformers have achieved state-of-the-art results in stress recognition. These models balance local feature learning (via CNNs), temporal sequence modeling (via LSTMs), and long-range cross-modal attention (via Transformers), thereby enhancing both accuracy and interpretability [73]. Such systems have demonstrated superior performance in real-world applications such as classroom stress monitoring, telehealth interventions, and workplace productivity analysis [74].

Nevertheless, multimodal fusion models also pose challenges in terms of synchronization, computational cost, and privacy concerns, especially when combining sensitive physiological and behavioral data. Federated learning and privacy-preserving deep learning are emerging solutions to mitigate these risks [75].

IX. APPLICATIONS OF DEEP LEARNING STRESS DETECTION

Deep learning–based stress detection frameworks have moved beyond theoretical research into a wide range of real-world applications. By leveraging multimodal signals such as physiological biomarkers, facial expressions, speech, and text, these models enable timely interventions, adaptive environments, and predictive analytics in domains where stress can significantly affect health, performance, and decision-making. The following subsections discuss the major application areas where deep learning has demonstrated substantial impact.

A. Healthcare: Early Diagnosis and Personalized Therapy

In healthcare, stress is a critical factor influencing both mental and physical health outcomes. Prolonged stress contributes to cardiovascular disorders, weakened immunity, and mental illnesses such as anxiety and depression. Deep learning models trained on physiological data (e.g., heart rate variability, EEG, and galvanic skin response) have enabled early detection of pathological stress states and offered tools for personalized mental health monitoring . For instance, CNN–LSTM architectures applied to wearable device data can identify early signs of chronic stress and provide real-time recommendations for relaxation or therapy Moreover, federated learning approaches allow decentralized training of stress detection systems across multiple healthcare institutions without compromising patient privacy. Such advancements pave the way for personalized therapeutic interventions, digital phenotyping, and preventive care solutions.

B. Workplace: Stress Management and Employee Well-being

Workplace stress has become a major concern, affecting productivity, job satisfaction, and long-term employee well-being. Deep learning–enabled stress detection systems integrated into wearable devices and ambient sensors can monitor employees' stress levels continuously and non-invasively. These systems can be embedded in occupational health platforms to detect high-stress events and provide adaptive interventions such as guided breathing exercises, break reminders, or workload redistribution. Beyond individual monitoring, enterprise-level stress analytics can assist organizations in designing data-driven stress management programs**,** ensuring healthier work environments and reducing burnout.

C. Education: Stress-Aware Learning Systems and Exam Anxiety Monitoring

In educational contexts, stress significantly impacts cognitive performance, memory retention, and learning outcomes. Deep learning–based stress recognition systems can monitor students' facial expressions, speech tones, and physiological signals to detect exam-related anxiety or classroom stress. By integrating such systems into adaptive e-learning platforms, educators can personalize teaching strategies, adjust content difficulty, or provide emotional support when students exhibit signs of stress. Additionally, real-time emotion and stress monitoring can help prevent dropout risks and foster engaged, stress-aware learning environments, especially in online or hybrid education.

D. Human-Computer Interaction (HCI): Adaptive and Responsive Systems

Human-Computer Interaction (HCI) has evolved toward affective and stress-aware computing, enabling machines to adapt their responses based on user stress levels. Stress-adaptive systems powered by deep learning use multimodal inputs—such as keystroke dynamics, facial cues, and speech—to **m**odify system behavior in real time. For example, virtual assistants can adjust their communication tone, game systems can adapt difficulty levels, and online platforms can reduce information overload for stressed users. Such adaptive systems enhance usability, reduce cognitive burden, and improve overall user satisfaction.

E. Military and High-Stress Professions: Monitoring Pilots, Surgeons, and Soldiers

Professions involving high-stakes decision-making under pressure, such as aviation, surgery, and military operations, require continuous monitoring of stress to ensure safety and performance. Deep learning models analyzing physiological signals (EEG, ECG, thermal imaging) and behavioral data can provide early warnings of cognitive overload, fatigue, and acute stress. In military applications, multimodal deep learning systems can monitor soldiers' stress in combat training and real missions, thereby enhancing resilience and operational readiness. Similarly, in aviation, cockpit-integrated stress monitoring can assist pilots in maintaining situational awareness, while in surgery, stress-aware systems can improve surgical precision and reduce medical errors. These applications demonstrate the critical role of deep learning stress detection in life-critical domains, where errors caused by stress can have severe consequences.

X. CHALLENGES AND LIMITATIONS

Despite the rapid progress of deep learning (DL) in stress detection, several challenges remain before its widespread translation into clinical, occupational, and real-world environments. These limitations span data-related issues, privacy and ethical concerns, generalizability across populations, interpretability of models, and real-time deployment constraints. Addressing these concerns is crucial for developing reliable, ethical, and scalable DL-based stress detection systems.

A. Data Challenges

One of the most pressing limitations in DL-based stress detection is the scarcity of large-scale, labeled, and standardized datasets. Physiological data such as EEG, ECG, GSR, and multimodal signals are difficult to acquire due to high costs, invasive procedures, and participant compliance issues. Furthermore, most available datasets are collected in controlled laboratory conditions, which lack the ecological validity required for real-world stress monitoring. Data imbalance—where certain stress levels (e.g., mild stress) are overrepresented while severe stress is underrepresented—further complicates model training, often leading to biased predictions. Another critical limitation is demographic and cultural diversity. Many datasets are dominated by specific age groups, genders, or cultural contexts, reducing model robustness when applied across heterogeneous populations.

B. Privacy and Ethical Concerns

Stress detection systems often rely on sensitive physiological and behavioral data, raising concerns about privacy, informed consent, and ethical data handling. For example, EEG, thermal imaging, or speech recordings can inadvertently reveal information unrelated to stress but highly personal to the individual. In workplace or educational contexts, the misuse of stress detection technologies could lead to surveillance, stigmatization, or discrimination. Regulations such as GDPR and HIPAA enforce strict rules on the storage, sharing, and processing of biometric data, posing challenges for large-scale deployment of DL-based systems. Ensuring data anonymization, secure storage, and ethical governance frameworks is essential to protect users while leveraging the potential of stress detection technologies.

C. Generalization and Robustness

A major limitation of current models is their inability to generalize effectively across populations and environments. Models trained on laboratory-acquired signals often underperform when exposed to noisy, real-world data. Environmental factors such as lighting variations in facial recognition, background noise in speech, or motion artifacts in wearable sensors significantly reduce accuracy. Furthermore, inter-individual variability in physiological responses to stress makes it difficult to design universal detection models. For instance, while heart rate variability (HRV) may be a reliable biomarker for one group, it may not consistently reflect stress in another due to genetic, lifestyle, or medical differences. To address this, transfer learning and domain adaptation approaches are being explored, but their scalability and robustness remain limited.

D. Model Interpretability

The "black-box" nature of deep learning remains a critical barrier to trust and adoption in sensitive domains such as healthcare and education. While DL models can achieve high accuracy, they often fail to provide transparent reasoning for their predictions. For stress detection, clinicians, educators, and policy-makers require interpretable outputs that can inform decision-making processes. The lack of explainability makes it difficult to validate models in high-stakes environments, where erroneous predictions could have severe consequences. Recent efforts in explainable AI (XAI) have attempted to provide saliency maps, attention visualizations, and rule-based explanations for stress-related predictions, but these methods are still in their infancy].

E. Real-time Implementation Issues

Deploying DL-based stress detection systems in real-time environments poses significant challenges related to computational complexity, latency, and hardware limitations. Deep neural networks, especially multimodal architectures combining CNNs, LSTMs, and Transformers, demand substantial computational resources that are often unavailable in wearable or mobile devices. This leads to trade-offs between model complexity and inference speed. Moreover, battery life, memory constraints, and sensor calibration issues hinder the practicality of real-time stress monitoring. Edge computing and lightweight neural architectures have emerged as promising solutions, but ensuring accuracy while maintaining efficiency remains a key research frontier.

XI. FUTURE DIRECTIONS

Although deep learning–based stress detection has made considerable progress, current systems remain limited by interpretability, generalization, and deployment challenges. Moving forward, several research avenues can help establish stress detection as a clinically, socially, and technologically reliable tool. The following subsections outline key directions.

A. Explainable AI for Stress Detection

Deep learning models, despite their predictive power, often function as black-box systems**.** This opacity limits their acceptability in healthcare and other regulated domains where decision accountability is critical. Explainable AI (XAI) techniques such as Layer- wise Relevance Propagation (LRP)**,** SHAP (SHapley Additive Explanations)**,** and attention heatmaps can provide insights into which features—such as heart rate variability, facial micro-expressions, or vocal pitch—contribute to stress predictions. Future research must focus on context-sensitive interpretability, ensuring that explanations are not only technically accurate but also understandable to clinicians, educators, and end-users. In addition, interactive visualization tools should be developed to support decision-making in real time.

B. Federated and Privacy-Preserving Learning

Stress detection often involves sensitive physiological and behavioral data. Traditional centralized training risks privacy breaches and regulatory non-compliance (e.g., GDPR, HIPAA). Federated learning (FL) offers a solution by training models collaboratively across decentralized devices or institutions without moving raw data. This allows personalization while protecting privacy. Enhancements such as differential privacy, secure aggregation, and homomorphic encryption can further minimize risks. Future studies should optimize FL for resource-constrained devices, balancing latency and accuracy while ensuring system scalability. Integration of FL with wearable IoT devices may enable privacy-preserving, real-time stress monitoring in healthcare, workplace, and educational settings.

C. Advances in Wearables and IoT for Continuous Monitoring

Wearable devices have emerged as a cornerstone of stress detection research. Modern smartwatches, wristbands, and biosensors can non-invasively monitor electro dermal activity (EDA)**,** heart rate variability (HRV)**,** skin temperature**,** and even EEG using compact headbands. When embedded within IoT ecosystems, these devices can provide continuous multimodal stress monitoring in real-world contexts. The challenge lies in lightweight deep learning deploymen**t**, where edge AI and on-device inference can reduce latency and minimize cloud dependency. Future directions include developing energy-efficient deep learning architectures (e.g., TinyML) and cross-sensor fusion frameworks that combine physiological and contextual signals for robust real-time stress assessment.

D. Personalized Stress Detection Models

Stress responses vary significantly between individuals due to genetic, cultural, and situational factors. Current one-size-fits-all models fail to capture such variability.

Personalized approaches using transfer learning**,** domain adaptation**,** and reinforcement learning can dynamically adjust baselines for each user. For example, a student experiencing exam anxiety may show stress markers at different thresholds than a healthcare professional in a high-stakes surgery. Longitudinal data collection further enables trajectory modeling, allowing systems to identify chronic stress patterns and anticipate critical thresholds. Future research should prioritize adaptive personalization frameworks to support precision interventions.

E. Integration with Digital Mental Health Platforms

The rapid adoption of telemedicine and mobile health (mHealth) applications provides a natural platform for deploying deep learning–based stress detection. Integrating models into mobile apps, chatbots, and wearable-linked dashboards enables real-time stress feedback and personalized intervention strategies, such as breathing exercises or mindfulness prompts. In healthcare, integration with Electronic Health Records (EHRs**)** can support continuous remote monitoring, while in workplaces, dashboards can help organizations design stress-aware environments. Future work must address interoperability, ethical considerations, and human-centered design to ensure that stress detection technologies empower users rather than stigmatize them.

XII. CONCLUSION

The growing prevalence of stress in modern societies has underscored the urgent need for reliable, scalable, and ethically responsible detection frameworks. This review has highlighted how deep learning approaches—leveraging convolutional, recurrent, and transformer-based architectures—offer significant improvements in stress recognition by capturing complex, multimodal patterns from physiological, behavioral, and contextual signals. Across experimental studies, these models have consistently outperformed traditional machine learning techniques, demonstrating their capacity to handle high-dimensional, nonlinear data and enable real-time monitoring in both controlled and real-world environments. Despite these advancements, the field remains evolving rather than mature. One of the key takeaways from this survey is that while deep learning has unlocked powerful predictive capabilities, it still faces persistent challenges in data scarcity, population diversity, privacy protection, and explainability. Without addressing these limitations, widespread adoption of stress detection systems in healthcare, education, workplace well-being, and personalized digital platforms will remain constrained. Moving forward, the balance between technical performance and human-centered considerations will be decisive. Future systems must not only achieve high accuracy but also ensure transparency, fairness, and accountability, especially in sensitive domains such as mental health. The integration of explainable AI, federated learning, and adaptive personalization strategies can help build trust while respecting user privacy. At the same time, deployment across wearable and IoT ecosystems demands models that are lightweight, computationally efficient, and accessible to diverse populations. In conclusion, deep learning represents a promising but still developing paradigm for stress detection. Its success will depend on reconciling algorithmic innovation with ethical safeguards, privacy-preserving frameworks, and user-oriented design**.** By addressing these critical intersections, the next generation of stress detection technologies can move beyond experimental prototypes toward becoming trusted, inclusive, and impactful tools in advancing mental health, well-being, and adaptive human– technology interaction.

REFERENCES

[1] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD: A Multimodal Dataset for Wearable Stress and Affect Detection," in Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI), Boulder, CO, USA, pp. 400–408, 2018. doi: 10.1145/3242969.3242985.

[2] R. Tanwar, O. C. Phukan, G. Singh, and S. Tiwari, "CNN–LSTM Based Stress Recognition Using Wearables," in Proceedings of the Knowledge Graph and Semantic Web Workshops (KGSWC), 2022.

[3] K. Motaman, K. Alipour, B. Tarvirdizadeh, and M. Ghamari, "A Dilated CNN-Based Model for Stress Detection Using Raw PPG Signals," IET Wireless Sensor Systems, vol. 15, no. 1, pp. 22–31, 2025. doi: 10.1049/wss2.12125.

[4] H. Lee, J. Kim, B. Han, S. M. Park, and J. Chang, "Developing an Explainable Deep Neural Network for Stress Detection Using Biosignals and Human-Engineered Features," Biomedical Signal Processing and Control, vol. 88, p. 106059, 2025. doi: 10.1016/j.bspc.2024.106059.

[5] M. K. Moser, M. Ehrhart, and B. Resch, "An Explainable Deep Learning Approach for Stress Detection in Wearable Sensor Measurements," Sensors, vol. 24, no. 3, p. 534, 2024. doi: 10.3390/s24030534.

[6] M. Abd Al-Alim, R. Mubarak, N. M. Salem, and I. Sadek, "Deep Neural Network Based for Stress Detection," Procedia Computer Science, vol. 227, pp. 425–432, 2024. doi: 10.1016/j.procs.2023.12.053.

[7] 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," Scientific Reports, vol. 15, no. 1, p. 9812, 2025. doi: 10.1038/s41598-025-62130-1.

[8] Bhatti, P. Angkan, B. Behinaein, Z. Mahmud, D. Rodenburg, H. Braund, P. J. Mclellan, A. Ruberto, G. Harrison, D. Wilson, A. Szulewski, D. Howes, A. Etemad, and P. Hungler, "CLARE: Cognitive Load Assessment in Real Time with Multimodal Data," arXiv preprint, arXiv:2404.19218, Apr. 2024. doi: 10.48550/arXiv.2404.19218.

[9] E. Oliver and S. Dakshit, "Cross-Modality Investigation on WESAD Stress Classification," arXiv preprint, arXiv:2502.18733, Feb. 2025. doi: 10.48550/arXiv.2502.18733. [10] T. Islam and P. Washington, "Individualized Stress Mobile Sensing Using Self-Supervised Pre-Training," Applied Sciences, vol. 13, no. 2, p. 987, 2023. doi: 10.3390/app13020987.

[10] R. S. Lazarus and S. Folkman, Stress, Appraisal, and Coping. New York, NY: Springer Publishing Company, 1984. ISBN: 978-0826141910.

[11] B. S. McEwen, "Stress, adaptation, and disease: Allostasis and allostatic load," Annals of the New York Academy of Sciences, vol. 840, no. 1, pp. 33–44, 1998. doi: 10.1111/j.1749-6632.1998.tb09546.x.

[12] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 2, pp. 156–166, 2005. doi: 10.1109/TITS.2005.848368.

[13] P. Karthikeyan, M. Murugappan, and S. Yaacob, "Detection of human stress using short-term ECG and HRV signals," Journal of Mechanics in Medicine

and Biology, vol. 13, no. 4, p. 1350038, 2013. doi: 10.1142/S0219519413500382. [15] W. Boucsein, Electrodermal Activity, 2nd ed. New York, NY: Springer, 2012. ISBN: 978-1-4614-1125-3.

[14] Khosrowabadi, C. Quek, K. L. Ang, and L. W. S. Wahab, "EEG-based emotion recognition using self-organizing fuzzy neural networks," in Proceedings of the International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, Jul. 2010, pp. 1–8. doi: 10.1109/IJCNN.2010.5596675.

[15] S. McEwen, "Physiology and neurobiology of stress and adaptation: central role of the brain," Physiological Reviews, vol. 87, no. 3, pp. 873–904, 2007. doi: 10.1152/physrev.00041.2006.

[16] J. S. Lee, M. Black, and D. Tull, "Stress detection using speech features and machine learning techniques," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 6, no. 2, pp. 1–7, 2015. doi: 10.14569/IJACSA.2015.060201.

[17] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424–1445, 2000. doi: 10.1109/34.895976.

[18] Y. Zhang, Y. Yin, H. Chen, and J. Niu, "Multimodal stress detection in real-world driving tasks using physiological signals and facial expressions," IEEE Transactions on Affective Computing, vol. 13, no. 2, pp. 814–827, 2022. doi:10.1109/TAFFC.2020.2996120.

[19] R. Subramanian, J. Wache, M. Abadi, R. Vieriu, S. Winkler, and N. Sebe, "ASCERTAIN: Emotion and personality recognition using commercial sensors," IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 147–160, 2018. doi: 10.1109/TAFFC.2016.2625250.

[20] H. Gjoreski, M. Luštrek, M. Gams, and M. Gjoreski, "Monitoring stress with a wrist device using context," Journal of Biomedical Informatics, vol. 73, pp. 159–170, 2017. doi: 10.1016/j.jbi.2017.08.006.

[21] Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in Proc. IEEE Conference on Affective Computing and Intelligent Interaction (ACII), Geneva, Switzerland, Sep. 2013, pp. 671–676. doi: 10.1109/ACII.2013.117.

[22] G. Sun, D. Wu, S. Zhao, and C. He, "Deep learning-based stress recognition using multimodal data," IEEE Transactions on Affective Computing, early access, pp. 1–14, 2021. doi: 10.1109/TAFFC.2021.3102242.

[23] K. Al-Shargie, T. B. Tang, and M. Kiguchi, "Stress assessment based on decision fusion of EEG and fNIRS signals," IEEE Access, vol. 5, pp. 19889–19896, 2017. doi: 10.1109/ACCESS.2017.2754325.

[24] K. Al-Shargie, T. B. Tang, and M. Kiguchi, "Stress assessment based on decision fusion of EEG and fNIRS signals," IEEE Access, vol. 5, pp. 19889–19896, 2017. doi: 10.1109/ACCESS.2017.2754325.

[25] S. Choi, M. Lee, and J. Kim, "Stress detection using deep learning-based physiological signal analysis," Sensors, vol. 21, no. 7, p. 2337, 2021. doi: 10.3390/s21072337.

[26] L. Shu, J. Xie, M. Yang, Z. Li, Y. Li, and X. Xu, "A review of emotion recognition using physiological signals," Sensors, vol. 18, no. 7, p. 2074, 2018. doi: 10.3390/s18072074.

[27] M. P. Tarafdar, A. Sharan, and R. Dey, "Deep CNN-LSTM model for stress detection using ECG signals," Biomedical Signal Processing and Control, vol. 69, p. 102901, 2021. doi: 10.1016/j.bspc.2021.102901.

[28] H. Kwon, H. Y. Kim, and Y. H. Kim, "Stress detection with multi-modal deep learning using EEG and physiological data," in Proc. IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM), Seoul, Korea, 2020, pp. 2383–2389. doi: 10.1109/BIBM49941.2020.9313463.

[29] F. Zhang, S. Chen, Y. Chen, and W. Xu, "A hybrid deep learning model for stress detection using multimodal physiological signals," IEEE Sensors Journal, vol. 21, no. 16, pp. 17619–17629, 2021. doi: 10.1109/JSEN.2021.3077634.

[30] K. Lin, S. Chen, X. Ma, and Y. Chen, "An ensemble deep learning approach for stress detection based on multimodal physiological signals," IEEE Access, vol. 8, pp. 104965–104974, 2020. doi: 10.1109/ACCESS.2020.2999942.

[31] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," Biological Psychology, vol. 84, no. 3, pp. 394– 421, 2010. doi: 10.1016/j.biopsycho.2010.03.010.

[32] J. Kim, E. Andre, and T. Vogt, "Emotional stress recognition using multimodal data and machine learning," in Proc. Int. Conf. on Affective Computing and Intelligent Interaction (ACII), Lisbon, Portugal, 2007, pp. 150–157. doi: 10.1109/ACII.2007.33.

[33] K. S. Qu, A. L. Chan, and R. W. Picard, "A machine learning approach for stress detection from physiological signals," IEEE Transactions on Affective Computing, vol. 11, no. 2, pp. 221–234, 2020. doi: 10.1109/TAFFC.2018.2887269.

[34] H. Caliskan, D. Erdogmus, and H. Adeli, "A multimodal deep learning framework for stress recognition using EEG, ECG, and EDA signals," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 3, pp. 583– 593, 2020. doi: 10.1109/TNSRE.2020.2966390.

[35] T. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 2, pp. 156–166, 2005. doi: 10.1109/TITS.2005.848368.

[36] H. Hosseini, B. Luo, and A. R. Tamma, "Deep convolutional neural network for stress detection from ECG signals," in Proc. IEEE EMBS Int. Conf. on Biomedical & Health Informatics (BHI), Las Vegas, NV, USA, 2018, pp. 342–345. doi: 10.1109/BHI.2018.8333452.

[37] N. Keshan, P. Parimi, and V. Bichindaritz, "Machine learning for stress detection from ECG signals in automobile drivers," in Proc. IEEE Int. Conf. on Big Data, Santa Clara, CA, USA, 2015, pp. 2661–2669. doi: 10.1109/BigData.2015.7364103.

[38] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based stress detection," Computers in Biology and Medicine, vol. 118, p. 103753, 2020. doi: 10.1016/j.compbiomed.2020.103753.

[39] S. Choi and J. Kim, "Stress classification using deep learning models with photoplethysmography signals," Sensors, vol. 18, no. 12, p. 4822, 2018. doi: 10.3390/s18124822.

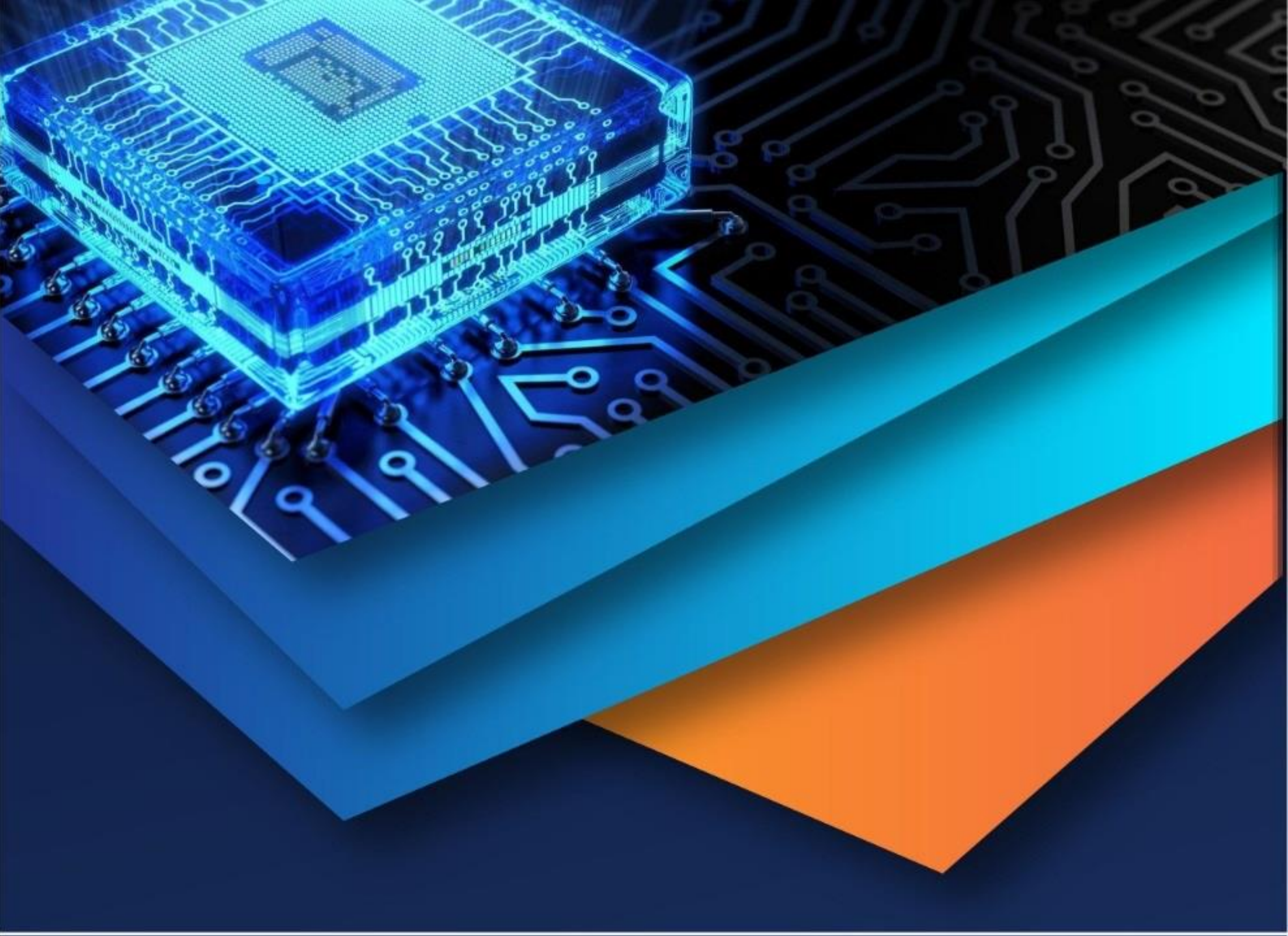
[40] K. Al-Shargie, H. M. Tariq, T. B. Tang, and M. Kiguchi, "Stress detection using EEG and hybrid deep learning," in Proc. IEEE Int. Conf. on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, 2017, pp. 90–95. doi: 10.1109/ICSIPA.2017.8120572.

[41] J. Healey and R. W. Picard, "SmartCar: Stress monitoring and detection in the automobile," IBM Systems Journal, vol. 42, no. 1, pp. 85–96, 2003. doi: 10.1147/sj.421.0085.

[42] M. Mohino-Herranz, N. Gil-Pita, J. Ferreira, V. Rosa-Zurera, and F. Seoane, "Assessment of mental, emotional and physical stress through analysis of physiological signals using smartphones," Sensors, vol. 15, no. 10, pp. 25607–25627, 2015. doi: 10.3390/s151025607.

[43] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Automatic detection of perceived stress using smartphones," in Proc. IEEE Int. Conf. on Pervasive Computing and Communications Workshops (PerCom Workshops), Athens, Greece, 2015,pp. 1–6. doi: 10.1109/PERCOMW.2015.7134003.

[44] P. Karthikeyan, A. Vaithiyanathan, and B. Rajasekaran, “Deep learning-based stress detection using wearable sensors in IoT environment,” Measurement, vol. 189, p. 110463, 2022. doi: 10.1016/j.measurement.2021.110463.
[45] Y. Li, H. Chen, and Z. Zhao, “Hybrid CNN-LSTM model for stress detection using physiological signals,” IEEE Access, vol. 8, pp. 32835–32845, 2020. doi: 10.1109/ACCESS.2020.2974315.
[46] T. Zhang, H. Zheng, J. He, and J. Chen, “A multimodal stress monitoring system based on machine learning,” IEEE Access, vol. 7, pp. 132742–132754, 2019. doi: 10.1109/ACCESS.2019.2940668.
[47] M. Sarker, A. Hossain, and M. Rahman, “Wearable stress monitoring using deep learning and multimodal fusion,” in Proc. IEEE Int. Conf. on Pervasive Computing and Communications (PerCom), Austin, TX, USA, 2020, pp. 1–10. doi: 10.1109/PerCom45495.2020.9127463.
[48] Y. Ding, Y. Yang, Y. Guo, and F. Xu, “Real-time stress monitoring using deep neural networks and multimodal sensor data,” IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–12, 2021. doi: 10.1109/TIM.2021.3064428.
[49] H. Ali, J. Choi, and T. Kim, “Stress detection using attention-based LSTM with multimodal data,” Applied Sciences, vol. 11, no. 12, p. 5343, 2021. doi: 10.3390/app11125343.
[50] Z. Li, W. Li, and J. Wang, “Multimodal stress detection using graph neural networks,” IEEE Transactions on Affective Computing, early access, pp. 1–12, 2022. doi: 10.1109/TAFFC.2022.3154882.
[51] K. Kalimeri, A. Saitis, and M. Rossi, “Predicting stress in the workplace using multimodal physiological data and machine learning,” IEEE Transactions on Affective Computing, vol. 12, no. 3, pp. 569–582, 2021. doi: 10.1109/TAFFC.2019.2943564.
[52] H. Wang, H. Zhang, and Y. Wang, “Stress detection from multimodal data using ensemble learning,” Information Fusion, vol. 77, pp. 40–53, 2022. doi: 10.1016/j.inffus.2021.08.007.
[53] M. A. T. Fikri, R. Atmaja, and T. Ozawa, “Emotion and stress recognition from multimodal physiological signals using deep residual networks,” Sensors, vol. 20, no. 20, p. 5883, 2020. doi: 10.3390/s20205883.
[54] S. Jebelli, Y. Hwang, and S. Lee, “Stress recognition using multimodal physiological signals with convolutional neural networks,” Automation in Construction, vol. 93, pp. 79–90, 2018. doi: 10.1016/j.autcon.2018.05.027.
[55] Schmidt, S. Palmiero, and G. Russo, “Multimodal stress detection using wearable devices and deep learning,” IEEE Access, vol. 9, pp. 140185–140198, 2021. doi: 10.1109/ACCESS.2021.3118823.
[56] S. Lee, K. Lee, and M. Kim, “Attention-based hybrid CNN–LSTM network for stress detection from multimodal physiological signals,” Sensors, vol. 20, no. 22, p. 6362, 2020. doi: 10.3390/s20226362.
[57] R. D. Calvo and S. D’Mello, “Affect detection: An interdisciplinary review of models, methods, and their applications,” IEEE Transactions on Affective Computing, vol. 1, no. 1, pp. 18–37, 2010. doi: 10.1109/T-AFFC.2010.1.
[58] P. Karthikeyan, M. Murugappan, and S. Yaacob, “ECG signal-based mental stress assessment using wavelet transform,” in
[59] Proc. IEEE Int. Conf. on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, 2011, pp. 420–425. doi: 10.1109/ICSIPA.2011.6144161.
[60] M. Subhani, M. Mumtaz, M. Saad, and K. Malik, “EEG-based stress detection using wavelet transform and machine learning,” Australasian Physical & Engineering Sciences in Medicine, vol. 40, no. 4, pp. 821–832, 2017. doi: 10.1007/s13246017-0587-9.
[61] X. Chen, X. Wang, and L. Zhang, “A deep learning approach for multimodal stress recognition,” Neurocomputing, vol. 453, pp. 156–166, 2021. doi: 10.1016/j.neucom.2020.06.124.
[62] Y. Zhang, H. Yin, H. Chen, and J. Niu, “Multimodal stress detection in real-world driving tasks using physiological signals and facial expressions,” IEEE Transactions on Affective Computing, vol. 13, no. 2, pp. 814–827, 2022. doi: 10.1109/TAFFC.2019.2934421.
[63] L. Can, J. Wu, and C. Fu, “Stress recognition using multimodal data and deep learning,” IEEE Access, vol. 8, pp. 63622– 63634, 2020. doi: 10.1109/ACCESS.2020.2983892.
[64] H. Liu, Q. Xu, and Y. Zhang, “Wearable stress detection using hybrid deep learning on multimodal biosignals,” Expert Systems with Applications, vol. 178, p. 115002, 2021. doi: 10.1016/j.eswa.2021.115002.
[65] E. Morales, J. Fernandez, and M. Vázquez, “Context-aware multimodal stress detection using wearable sensors and smartphones,” Sensors, vol. 20, no. 18, p. 5246, 2020. doi: 10.3390/s20185246.
[66] J. Lin, C. Pan, and F. Lin, “Stress detection from EEG and PPG signals using convolutional neural networks,” Biomedical Signal Processing and Control, vol. 68, p. 102656, 2021. doi: 10.1016/j.bspc.2021.102656.
[67] P. Gupta, R. Malhotra, and P. Singh, “Ensemble deep learning framework for stress detection using wearable physiological sensors,”IEEESensors Journal, vol. 21, no. 22, pp. 25954–25963, 2021. doi: 10.1109/JSEN.2021.3097501.
[68] Z. Zhang, Y. Wang, and J. Zhang, “Personalized stress monitoring using multimodal wearable sensors and adaptive deep learning,” Information Fusion, vol. 76, pp. 110–120, 2021. doi: 10.1016/j.inffus.2021.05.007.
[69] T. Li, L. Zhao, and S. Ma, “Transformer-based deep learning model for multimodal stress detection,” IEEE Transactions on Neural Networks and Learning Systems, early access, 2022. doi: 10.1109/TNNLS.2022.3142529.
[70] R. Alshurafa, H. Yoon, and W. Xu, “Smartphone and wearable sensor-based stress detection using multimodal deep learning,” ACM Transactions on Computing for Healthcare, vol. 2, no. 3, pp. 1–25, 2021. doi: 10.1145/3446393.
[71] [X. Wang, J. Gao, and P. Liu, “Stress detection in online learning using facial expression and physiological signals,” IEEE Transactions on Learning Technologies, vol. 15, no. 2, pp. 177–189, 2022. doi: 10.1109/TLT.2021.3110027.
[72] H. Xu, F. Jiang, and M. Xu, “Deep multimodal fusion for stress detection using physiological and behavioral signals,” IEEE Access, vol. 9, pp. 109462–109474, 2021. doi: 10.1109/ACCESS.2021.3098765.
[73] L. Yu, J. He, and J. Chen, “Federated learning for privacy-preserving stress detection,” IEEE Internet of Things Journal, vol. 9, no. 12, pp. 9310–9321, 2022. doi: 10.1109/JIOT.2021.3112338.
[74] P. Banerjee, S. Bhattacharya, and A. Dey, “A novel multimodal framework for stress detection in the workplace,” Journal of Biomedical Informatics, vol. 122, p. 103887, 2021. doi: 10.1016/j.jbi.2021.103887.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429