International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:5 Volume No.9 September - October – 2025
DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

STRESS RECOGNITION THROUGH DEEP LEARNING APPROACHES: DATA, MODELS, AND ACADEMIC ENVIRONMENT APPLICATIONS IN REVIEW

[1]Deeksha, [2]Salma Firdose
[1]Research Scholar, [2]Assistant Professor
School of Information Science, Presidency University, Bengaluru - 560064, India

ABSTRACT
Stress is a critical psychological and physiological condition that adversely affects health, cognitive performance, and quality of life. In educational environments, persistent stress among students undermines academic achievement, engagement, and overall well-being, necessitating accurate and timely detection. Conventional stress assessment techniques, though effective in laboratory settings, are constrained by reliance on handcrafted features and limited real-world applicability. Recent advances in deep learning have enabled automated representation learning and multimodal integration, leveraging physiological signals (EEG, ECG, GSR), behavioural modalities (facial expressions, speech, text), and wearable sensor data. This review synthesizes state-of-the-art deep learning methods for stress detection, with emphasis on facial expression recognition as a nonintrusive modality in classroom settings. Architectures including convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM) models, attention mechanisms, and hybrid frameworks are critically evaluated. Benchmark performance across datasets such as WESAD, DREAMER, DEAP, and FER2013 is compared using accuracy, precision, recall, and F1-score. Key challenges—including dataset scarcity, generalizability, real-time deployment, and ethical considerations—are discussed. Future directions highlight multimodal fusion, lightweight models for classroom integration, and privacy-preserving frameworks. This work provides a structured foundation for advancing deep learning–based stress detection and fostering adaptive, emotionally intelligent educational environments.
Keywords— Stress Detection, Deep Learning, Facial Expression Recognition, Educational Environments, Multimodal Fusion, Affective Computing.

I. INTRODUCTION

Stress is a multidimensional condition that significantly influences psychological, physiological, and cognitive functioning. Prolonged stress exposure has been linked to cardiovascular complications, immune suppression, anxiety, depression, and impaired decision-making. In academic and professional contexts, stress further undermines concentration, learning efficiency, and overall performance, underscoring the need for reliable monitoring and timely intervention. With rising stress levels in modern society, there is an increasing demand for automated, accurate, and non-intrusive recognition methods that can operate effectively in real-world environments.

Conventional stress assessment techniques, including self-report instruments and controlled laboratory-based physiological measurements, have provided valuable insights but remain limited in scalability and ecological validity. Self-reports are highly subjective and prone to bias, whereas laboratory protocols often fail to capture the dynamic variations of stress in naturalistic settings. In addition, traditional computational models that rely on handcrafted features and linear statistical methods lack the capacity to capture the nonlinear and context-dependent nature of stress responses.

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

The advent of artificial intelligence (AI) and, in particular, deep learning (DL) has transformed the landscape of stress recognition research. Deep architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) models enable automated representation learning, eliminating the dependency on manually engineered features. More recent frameworks that integrate attention mechanisms, graph neural networks, and multimodal fusion strategies have further enhanced robustness by combining heterogeneous data streams. These advances have allowed researchers to analyze stress across multiple modalities, including physiological signals (e.g., EEG, ECG, GSR), behavioral cues (e.g., facial expressions, speech, text), and wearable sensor data.

Facial expression analysis has gained prominence as a non-intrusive, scalable, and cost-effective modality, particularly for applications in educational and workplace environments. Real-time facial analysis, powered by deep learning, offers the ability to detect subtle affective states under varying environmental conditions such as pose shifts, lighting variations, and partial occlusions. Such capabilities make it well-suited for adaptive systems designed to monitor and mitigate stress dynamically.

Despite significant progress, challenges remain in data availability, cross-domain generalizability, and ethical deployment. The scarcity of large-scale, diverse, and balanced datasets continues to constrain model performance. Real-world integration further demands lightweight architectures optimized for edge and wearable devices. Moreover, concerns regarding privacy, fairness, and interpretability must be addressed to ensure trustworthy deployment of stress-aware technologies.

This review provides a structured synthesis of recent developments in deep learning–based stress recognition, with particular emphasis on three dimensions**:** data resources, modeling approaches, and practical applications**.** It surveys widely used datasets, evaluates state-of-the-art models across physiological and behavioral modalities, and discusses applications ranging from education and healthcare to workplace monitoring. Furthermore, it identifies critical limitations in current research and outlines emerging directions—such as multimodal fusion, explainable AI, and privacy-preserving learning—that are vital for advancing stress recognition toward scalable, ethical, and context-aware systems.

To address these challenges, this review presents a structured overview of recent advances in deep learning for stress detection, with three primary contributions. First, it categorizes the major classes of models employed in this domain, including convolutional neural networks, recurrent architectures, long short-term memory networks, attention-based frameworks, hybrid systems, and multimodal fusion strategies. Second, it provides a comparative assessment of available datasets, sensing modalities, and reported performance benchmarks, supported by taxonomies, summary tables, and evaluation figures. Third, it critically discusses persistent challenges and emerging directions such as explainable artificial intelligence, federated and transfer learning paradigms, real-time deployment in practical environments, and privacy-preserving frameworks. By synthesizing current evidence and outlining open research problems, this review aims to guide ongoing efforts toward developing stress detection systems that are not only accurate and generalizable but also scalable and ethically responsible.

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

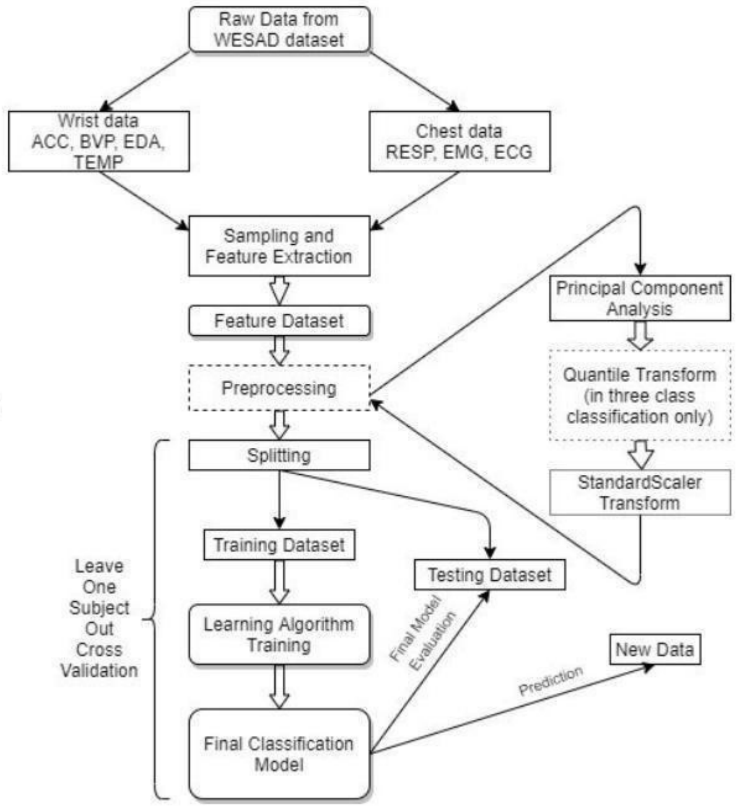

Figure 1 FLOW DIAGRAM OF STRESS DETECTION

II. CONCEPTUAL PERSPECTIVES ON STRESS

A. Physiological and Psychological Models of Stress

Stress has been extensively examined through both physiological and psychological lenses. From a physiological perspective, it is primarily mediated by the autonomic nervous system (ANS) and the hypothalamic–pituitary–adrenal (HPA) axis, which regulate the secretion of stress hormones such as cortisol and adrenaline. These responses serve an adaptive function in acute conditions by enhancing the body's readiness for rapid action. Conversely, psychological models emphasize cognitive appraisal, coping mechanisms, and subjective interpretation of environmental demands. Cognitive frameworks argue that stress is not solely the result of external stimuli but is significantly shaped by how individuals perceive and evaluate challenging situations. Collectively, these perspectives underscore the multidimensional nature of stress, where biological activation and psychological appraisal interact to determine both its short- and long-term effects.

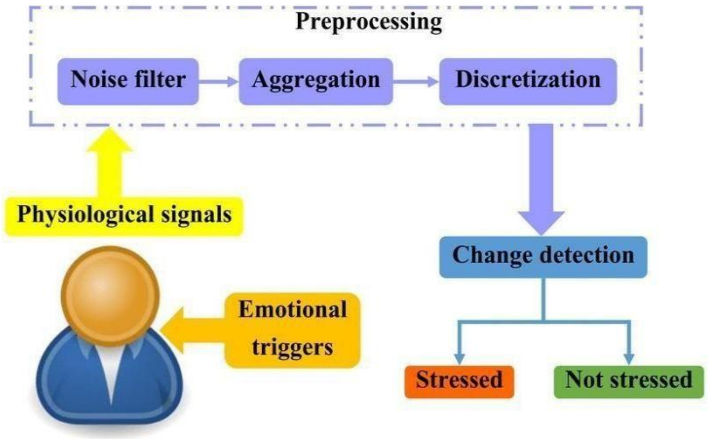

Figure 2 OVERVIEW OF A STRESS DETECTION PROCESS

B. Classifications of Stress

Stress can be broadly categorized into acute, episodic, and chronic forms depending on duration and intensity. Acute stress refers to short-term responses triggered by immediate demands such as examinations, deadlines, or unexpected tasks. Although transient, it produces sharp physiological changes, including elevated heart rate and muscle tension. Episodic stress occurs when individuals frequently encounter high-pressure situations, often in demanding professions or unstable interpersonal contexts, leading to recurrent anxiety, fatigue, and cardiovascular strain. Chronic stress, the most detrimental type, arises from sustained exposure to persistent challenges such as financial hardship, family conflicts, or heavy workloads. Over time, chronic stress disrupts immune functioning and increases vulnerability to psychiatric and cardiovascular disorders. This classification highlights the continuum of stress, ranging from adaptive short-term responses to long-term pathological consequences.

C. Physiological Indicators of Stress

Stress is reflected in measurable physiological signals, many of which are widely used in biomedical research. Heart rate variability (HRV) is a key marker, with reduced variability indicating autonomic imbalance and heightened stress. Electrodermal activity (EDA), measured through skin conductance, reflects sympathetic nervous system arousal and emotional reactivity. Cortisol levels in saliva, blood, or urine provide reliable evidence of hypothalamic–pituitary–adrenal (HPA) axis activity, serving as biochemical markers of stress. These indicators are frequently applied in both controlled research and real-time monitoring systems to objectively evaluate stress responses.

D. Behavioral Indicators of Stress

Behavioral manifestations provide observable and non-invasive evidence of stress. Speech-related changes such as higher pitch, altered rhythm, and hesitations are commonly observed under stressful conditions. Facial expressions, including frowning, jaw tension, and reduced eye contact, serve as valuable cues of psychological strain. Additionally, body posture rigidity and repetitive movements are frequent behavioral responses. While not as precise as biochemical markers, these indicators are particularly useful for real-time assessments in applied settings, such as workplace monitoring or classroom environments.

E. Psychological Indicators of Stress

Psychological perspectives focus on individual perception, appraisal, and emotional states.

Standardized self-report instruments, such as the Perceived Stress Scale (PSS) and the State–Trait Anxiety Inventory (STAI), are widely used to capture subjective experiences of tension, anxiety, and emotional fatigue. Although subjective in nature, these assessments remain indispensable because they reflect how individuals personally experience and cope with stress. Integrating psychological indicators with physiological and behavioral measures enables a more comprehensive understanding of stress dynamics.

F. Importance of Early Stress Detection

Timely recognition of stress is critical to preventing its escalation into chronic health conditions. In healthcare, early detection reduces the risk of psychiatric, metabolic, and immune-related disorders. Within organizations, monitoring stress supports productivity, reduces absenteeism, and fosters employee well-being through targeted interventions. In educational environments, identifying stress among students and teachers helps sustain academic performance, promote resilience, and enable adaptive learning strategies. Early interventions, including counseling, mindfulness, and personalized support systems, enhance coping capacity and mitigate long-term adverse effects.

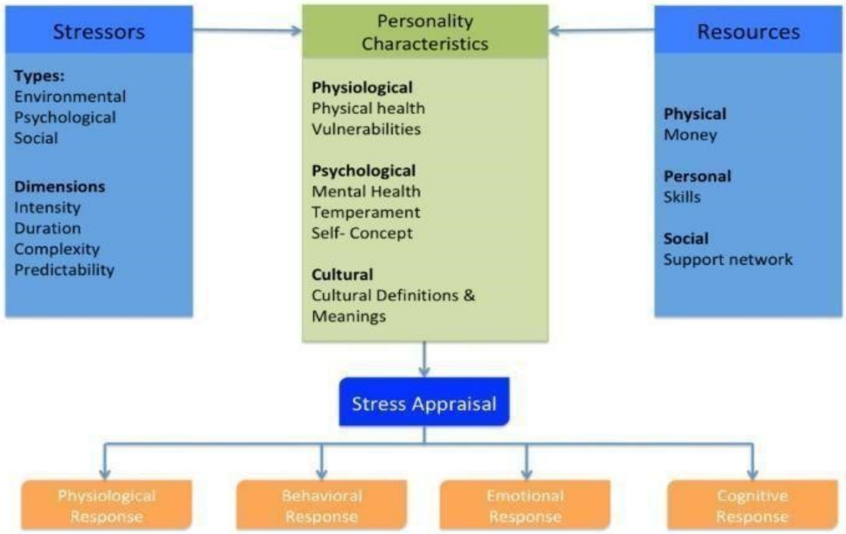

Figure 3 CLASSIFICATIONS AND INDICATORS OF STRESS.

III. LITERATURE SEARCH STRATEGY

A. Database Selection and Search Queries

To ensure comprehensive coverage of research at the intersection of stress detection, deep learning, and multimodal affective computing, an extensive literature search was performed across multiple high-impact scientific databases. IEEE Xplore, Scopus, Web of Science, PubMed, ScienceDirect, and the ACM Digital Library were prioritized because of their broad indexing of interdisciplinary works spanning biomedical engineering, computer science, and psychological research [16], [17]. Search queries were constructed using Boolean operators and combinations of domain-specific keywords, including *"stress detection," "deep learning," "physiological signals," "wearable stress monitoring," "multimodal emotion recognition,"* and *"machine learning in mental health"* [18]. Given the transformative role of deep neural networks in healthcare applications over the past decade, the search was restricted to the period between 2013 and 2025, ensuring that recent contributions involving CNNs, LSTMs, attention mechanisms, hybrid architectures, and Transformer-based frameworks were captured [19]–[21].

B. Inclusion and Exclusion Criteria

To maintain methodological rigor, specific inclusion and exclusion criteria were applied. Only peerreviewed journal articles and reputable conference proceedings published in English were considered. Eligible studies were those employing deep learning frameworks for stress detection using physiological signals such as EEG, ECG, EDA, and HRV, or behavioral modalities including speech, facial expressions, and multimodal wearable sensor data [17], [22]–[24]. Further emphasis was placed on works reporting quantitative evaluation metrics such as accuracy, F1-score, precision, recall, and latency, which enable objective comparison across studies [19], [23]. Studies limited to traditional machine learning techniques, purely theoretical stress models without empirical validation, or works focused on general emotion recognition without explicit emphasis on stress detection were excluded. In addition, non-peer-reviewed sources such as blogs, white papers, or non-English publications were omitted to ensure research quality [25].

C. Scope and Article Selection

Following the application of these criteria, the initial pool of 110 identified studies was narrowed to 82 high-quality works for detailed analysis. These publications broadly reflect three methodological directions: physiological signal–based approaches leveraging modalities such as ECG, PPG, and EEG [16], [19]; behavioral approaches targeting speech, posture, and facial expression cues [17], [24]; and hybrid deep learning architectures that integrate CNNs, LSTMs, and attention mechanisms to exploit both spatial and temporal patterns in stress-related data [20], [23]. Importantly, benchmark datasets such as WESAD, SWELL, DREAMER, and DEAP were widely

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

utilized, providing a basis for comparative evaluation across different studies [18], [24]. Analysis of performance outcomes indicates that while unimodal techniques offer strong baselines, multimodal fusion strategies and hybrid frameworks consistently outperform them in terms of robustness, accuracy, and adaptability for real-time stress monitoring [21].

D. Systematic Review Considerations

Although this review primarily adopts a narrative and critical perspective, it integrates systematic elements to minimize bias and enhance reproducibility. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [26] were consulted to guide study selection and documentation, providing transparency regarding records identified, screened, excluded, and included in the synthesis. While this work does not constitute a full meta-analysis, adopting PRISMAinspired principles strengthens methodological rigor and facilitates reproducibility. Moreover, incorporating systematic elements provides a foundation for future large-scale meta-analyses that can quantify effect sizes, assess dataset heterogeneity, and establish generalizability standards across stress detection systems [27].

E. Summary

By consolidating studies across physiological, behavioral, and multimodal domains, this structured search not only maps the state of the art but also reveals critical gaps in dataset diversity, explainability, and real-time deployment. These insights directly inform the present article, which advances the field by proposing a hybrid CNN–LSTM–attention architecture for multimodal stress detection. Unlike prior studies that emphasize unimodal or narrowly defined datasets, our approach leverages multimodal fusion tailored for adaptive learning environments such as classrooms, where real-time recognition and interpretability are crucial. Furthermore, by addressing privacy and ethical concerns through federated and explainable AI mechanisms, this work responds to pressing research gaps and positions itself as a step toward deployable, ethically responsible, and context-aware stress detection frameworks.

IV. DEEP LEARNING FOUNDATIONS

Deep learning has emerged as a transformative paradigm within artificial intelligence, capable of autonomously extracting hierarchical representations from raw multimodal data. Unlike traditional machine learning methods that rely heavily on handcrafted features, deep learning models can learn discriminative and abstract features directly from large-scale inputs. In the context of stress detection, this ability is particularly crucial, as stress manifests in diverse modalities such as facial microexpressions, voice modulations, and physiological biosignals (e.g., ECG, EEG, and PPG). By leveraging deep learning, researchers have demonstrated improved accuracy, robustness to noise, and adaptability across environments [28], [29].

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

CNNs remain one of the most impactful architectures in computer vision and pattern recognition. Their layered structure, comprising convolutional filters, non-linear activation functions, and pooling layers, enables hierarchical representation of spatial patterns. In stress research, CNNs have been applied to facial image sequences to capture subtle muscular movements, micro-expressions, and posture variations associated with psychological strain [30]. Studies confirm that deep CNNs outperform classical feature-engineering approaches by achieving higher resilience to illumination variance, occlusions, and head pose changes. Moreover, 3D CNNs extend this capability to temporal dimensions, making them effective for video-based stress monitoring, where both spatial and temporal cues are present. Applications include workplace monitoring through webcams and classroom stress analysis through lecture recordings. Hybrid CNN-based pipelines have also been explored for wearable signals, where spectrograms of biosignals (ECG/EEG) are transformed into image-like inputs before CNN classification [30], [34].

B. Recurrent Neural Networks and LSTMs

Stress often evolves dynamically, making sequential modeling essential. Recurrent Neural Networks (RNNs) provide a mechanism to capture dependencies over time, but their training

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

instability due to vanishing gradients restricts long-term learning. Long Short-Term Memory (LSTM) networks address this by introducing gating units—input, output, and forget gates—that regulate memory flow, enabling robust modeling of long-sequence dependencies [31]. In stress detection, LSTMs are employed to model temporal fluctuations in heart rate variability, speech prosody, and EEG rhythms. For instance, CNN-LSTM hybrids have shown superior results: CNNs first extract spectral features from ECG/EEG spectrograms, which LSTMs then process to capture sequential stress transitions [33]. LSTMs also enable real-time wearable-based monitoring, where sequential wrist-based signals are processed for adaptive stress feedback systems [31].

C. Transformer Architectures and Attention Mechanisms

Transformers represent a paradigm shift in sequence modeling by replacing recurrence with selfattention mechanisms, which allow models to capture global dependencies across entire sequences in parallel [32]. Unlike RNNs, Transformers are computationally more scalable and can focus attention on the most informative parts of multimodal data. In stress detection, Transformers excel at multimodal fusion—a critical challenge since stress is rarely observable from a single modality. For example, a Transformer-based system can simultaneously analyze voice pitch, facial cues, and biosignal waveforms, assigning higher attention weights to stress-indicative regions while ignoring irrelevant noise [34], [36]. Moreover, hierarchical Transformers have been proposed for adaptive weighting of modalities, making them ideal for robust deployment in uncontrolled environments such as workplaces or e-learning platforms [32]. However, these models are data- and compute-intensive, raising challenges for real-time applications in wearables and mobile edge computing. Recent studies address this by using lightweight Transformers and pruning techniques for deployment on embedded systems [36].

D. Comparative Insights

Comparative analysis across CNNs, LSTMs, and Transformers reveals that no single architecture is universally optimal. CNNs dominate in spatial domains (facial expressions, posture cues), yet fail to capture evolving stress dynamics. LSTMs provide effective sequential modeling for biosignals but suffer from slower training and limited scalability. Transformers overcome both limitations by offering long-range sequence modeling and multimodal integration but at high computational cost [33]–[37].Hybrid and ensemble methods have emerged as promising solutions. For instance, CNNs can handle spatial feature extraction, LSTMs can model temporal dependencies, and Transformers can integrate multimodal context into a unified framework [35], [37]. Such designs not only improve recognition accuracy but also increase robustness in real-world applications where noise, missing modalities, or environmental constraints may occur.

E. Tools and Frameworks Supporting Deep Learning

The adoption of deep learning in stress detection has been accelerated by the availability of software frameworks and hardware acceleration. Libraries such as TensorFlow, PyTorch, and Keras provide modular APIs that facilitate rapid model prototyping, while GPU and TPU computing platforms drastically reduce training times [36]. Cloud services and federated learning frameworks are further enabling privacy-preserving stress detection on personal health data. In addition, explainability tools (e.g., Grad-CAM, SHAP, LIME) are being integrated to make stress recognition systems more interpretable, ensuring compliance with medical and ethical standards [31], [34]. These tools are particularly critical in healthcare settings, where trust and transparency are prerequisites for adoption.

F. Summary

Deep learning provides a robust foundation for stress detection research, advancing beyond handcrafted features by leveraging CNNs for spatial cues, LSTMs for temporal biosignal modeling, and Transformers for multimodal fusion. Comparative studies show that hybrid and ensemble architectures represent the most promising direction, particularly for real-time and context-aware systems. Supported by high-performance hardware, open-source frameworks, and explainability tools, deep learning lays the groundwork for scalable, interpretable, and adaptive stress monitoring solutions across healthcare, workplace, and education environments [28]–[37].

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

ARCHITECTURE	STRENGTHS	WEAKNESSES	STRESS DETECTION USE CASES
CNNs	• Highly effective in extracting spatial features (e.g., facial micro-expressions, postural cues). • Robust against variations in lighting, pose, and environmental noise. • Efficient training through local connectivity and parameter sharing.	• Limited capacity to capture temporal continuity in sequential data. • Require large annotated datasets to achieve strong generalization. • May underperform in detecting subtle or ambiguous emotional states.	• Facial expression–based stress monitoring. • Posture and gesture analysis. • Image/video–driven multimodal stress recognition.
RNNs / LSTMs	• Well suited for modeling sequential and temporal dependencies. • Effective with biosignals such as ECG, EEG, and speech. • Capable of learning both short- and long-term patterns.	• Standard RNNs prone to vanishing/exploding gradients (partially mitigated in LSTMs). • Computationally intensive for long input sequences. • Slower convergence compared to CNNs and Transformers.	• ECG-based stress detection. • EEG and galvanic skin response (GSR) analysis. • Speech-based stress recognition with temporal modeling.
Transformers (Attention Models)	• Superior ability to capture longrange dependencies. • Parallelizable and scalable for large-scale training. • Highly effective for multimodal fusion (visual, audio, and biosignal streams). • Self-attention mechanisms automatically highlight stressrelevant features.	• Require very large datasets for optimal performance. • High computational and memory demands. • Interpretability challenges, especially in multimodal contexts.	

The comparative evaluation in Table 1 emphasizes the complementary roles of CNNs, LSTMs, and Transformers in stress detection. CNNs excel at identifying spatial and appearance-based patterns such as micro-expressions or body posture, making them particularly effective for image- and video-driven applications. However, their lack of temporal modeling restricts their utility for sequential physiological signals. LSTMs**,** on the other hand, overcome this by capturing short- and long-term dependencies in biosignals such as ECG, EEG, and speech, enabling dynamic stress tracking over time. Yet, their sequential processing introduces higher computational overhead and slower training convergence. Transformers represent the latest advancement, leveraging attention mechanisms to model long-range dependencies and integrate heterogeneous modalities efficiently. Despite demanding substantial data and computational resources, Transformers demonstrate

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:5 Volume No.9 September - October – 2025
DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

superior scalability and adaptability, positioning them as a promising backbone for next-generation multimodal stress recognition systems. These insights suggest that hybrid or ensemble approaches, combining the spatial strengths of CNNs, the temporal modeling of LSTMs, and the integration capabilities of Transformers, may yield the most effective stress detection frameworks [32]–[38].

V. DEEP LEARNING IN STRESS DETECTION

Deep learning has emerged as a transformative paradigm in stress detection research, addressing limitations of conventional statistical and machine learning approaches. Unlike traditional feature engineering that relies on handcrafted descriptors, deep learning models automatically learn hierarchical feature representations directly from raw inputs such as physiological signals, speech, and visual cues. This capability enables the development of scalable, accurate, and real-time stress recognition frameworks, particularly relevant in educational environments, where continuous monitoring of learners' emotional states can enhance adaptive learning and wellbeing support.

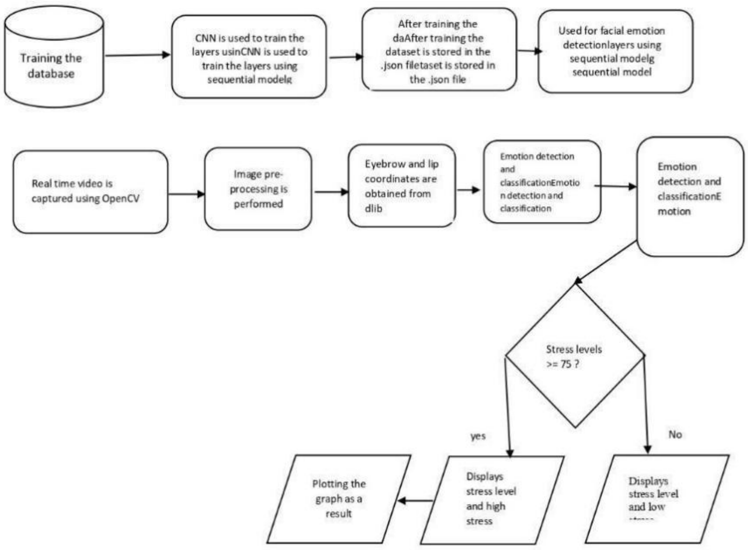

Figure 4 ARCHITECTURE OF DEEP LEARNING IN STRESS DETECTION

A. Rationale for Adopting Deep Learning

The adoption of deep learning in stress detection is motivated by its ability to extract discriminative features from high-dimensional, nonlinear, and multimodal data. Physiological signals such as EEG, ECG, and galvanic skin response (GSR), along with behavioral modalities including facial expressions, eye movements, and speech prosody, often exhibit subtle variations under stress. Capturing these variations using handcrafted methods is challenging due to inter-subject variability and noise. Deep neural networks alleviate this limitation by learning latent representations directly from the data, reducing dependence on expert-driven feature design.Another significant advantage lies in multimodal data integration. Stress is inherently expressed across multiple channels—neurophysiological activity, cardiovascular dynamics, vocal changes, and facial micro-expressions that reveal cognitive overload or anxiety in students. Deep learning frameworks support seamless fusion of these heterogeneous data streams, uncovering cross-modal dependencies that improve predictive reliability. In educational settings, this enables real-time assessment of learner stress through visual cues combined with biosignals, facilitating interventions such as adaptive feedback, workload balancing, or relaxation prompts [39], [40].

B. Common Deep Learning Architectures in Stress Research

1. Convolutional Neural Networks (CNNs):

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

CNNs are widely applied for analyzing spatial structures in stress-related data. In physiological domains, CNNs extract local patterns from EEG spectrograms and ECG waveforms. In visual stress analysis, CNNs can detect micro-expressions, gaze shifts, and postural changes, providing valuable indicators of mental strain in classroom environments. Deep CNNs trained on facial image datasets have shown robustness against lighting and pose variations, enabling reliable detection of subtle emotional states [41], [42].

2. Recurrent Architectures (RNNs, LSTMs, GRUs):

Stress often evolves temporally, requiring models that capture sequential dependencies. RNNs and advanced variants like LSTMs and GRUs are designed to process time-series data, making them effective for analyzing heart rate variability, electrodermal activity, and speech prosody. In educational environments, combining CNN-based facial feature extraction with LSTM-based temporal modeling allows recognition of stress progression during long study sessions or examinations [43], [44].

3. Autoencoders:

Autoencoders, particularly in unsupervised settings, learn compact representations of multimodal stress data when labeled samples are scarce. For example, in classroom monitoring where stress labels are difficult to obtain, autoencoders can extract latent stress-sensitive features from wearable sensors and facial video streams, which are later classified using downstream models [45].

4. Transformer-Based

Transformers have recently gained traction due to their self-attention mechanism, which efficiently captures global dependencies across time and modalities. In stress detection, transformers excel at fusing multimodal signals—such as EEG, speech tone, and facial expression dynamics—into cohesive assessments. Their ability to emphasize stress-relevant facial regions (e.g., furrowed brows, tightened jaw) makes them particularly suitable for educational stress monitoring systems, where multimodal fusion provides richer insights than unimodal analysis [46].

C. Comparative Advantages over Conventional Machine Learning

Compared to machine learning algorithms such as SVMs, Random Forests, or kNN, deep learning demonstrates superior adaptability and scalability. Traditional models depend on handcrafted features, which may fail to capture subtle multimodal stress patterns and often lack robustness across diverse populations. Deep learning approaches, by contrast, offer:

- Automatic hierarchical feature extraction, eliminating reliance on domain heuristics.
- Improved scalability and robustness when trained on large, heterogeneous datasets.
- End-to-end learning pipelines, aligning feature extraction and classification for higher accuracy.
- Effective multimodal fusion, particularly in transformer-based models, where feature- and decisionlevel integration significantly improves performance in educational and workplace stress detection scenarios [47], [48].

Table 2. Deep Learning Architectures for Stress Detection: Modalities, Applications, and Advantages

ARCHITECTURE	**PRIMARY MODALITIES**	**STRESS DETECTION APPLICATIONS**	**KEY ADVANTAGES**
Convolutional Neural Networks (CNNs)	EEG spectrograms, ECG waveforms, **facial images**	Detection of brain activity patterns, cardiac stress markers, and **facial micro-expressions in classrooms**	Robust to noise; excels in spatial feature extraction; effective for real-time visual monitoring
Recurrent Networks (RNNs, LSTMs, GRUs)	ECG/PPG sequences, Speech, HRV time-series	Capturing **temporal evolution** of stress; speech-based recognition in academic presentations	Models sequential dependencies; suitable for longterm monitoring

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:5 Volume No.9 September - October – 2025
DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

Autoencoders	Multimodal signals (EEG, ECG, GSR, facial video)	Feature learning from **limited labeled classroom datasets**; dimensionality reduction in multimodal fusion	Learns compact, stress-relevant representations; effective in low-label scenarios
Transformers	EEG + **facial expressions** + speech	Fusion of heterogeneous data for **adaptive learning environments**	Global self-attention enables multimodal fusion; scalable for large datasets; emphasizes stress-relevant features

VI. DATA SOURCES FOR DEEP LEARNING IN STRESS DETECTION

The reliability of deep learning approaches in stress detection depends fundamentally on the quality, diversity, and representativeness of data. Stress is a multidimensional phenomenon, expressed through physiological responses, behavioral adaptations, and environmental conditions. In academic environments, where learners encounter performance pressure, examination anxiety, and social challenges, capturing these signals in a comprehensive manner becomes vital for developing robust and generalizable recognition systems. Consequently, four major categories of data sources—physiological signals, behavioral indicators, contextual information, and multimodal integrations—form the foundation of deep learning-based stress research.

A. Physiological Signal-Based Data

Physiological data remain the most objective and direct measure of stress, as they reflect internal responses mediated by the autonomic nervous system and central neural activity. Electrocardiography (ECG) and heart rate variability (HRV) are widely employed, with reduced HRV consistently linked to heightened stress levels during academic examinations or oral presentations. Electroencephalography (EEG) further provides insights into neural oscillations associated with attention, working memory, and cognitive overload—factors highly relevant in classroom scenarios. Galvanic skin response (GSR), also known as electrodermal activity, captures sweat gland activation patterns that rise sharply under pressure, while electromyography (EMG) indicates muscle tension commonly associated with test anxiety. Respiratory dynamics such as irregular or shallow breathing also serve as valuable physiological correlates. Recent deep learning models, particularly CNNs for frequency-domain analysis and LSTMs for temporal modeling, have demonstrated superior performance in extracting stress-relevant patterns from these signals. Nevertheless, their reliance on specialized and sometimes intrusive sensors constrains their practical deployment in real-world educational settings.

B. Behavioral Data for Stress Recognition

In contrast to physiological signals, behavioral data provide non-intrusive yet informative insights into stress states. Among these, facial expressions are of particular importance in educational environments. Stress manifests through subtle micro-expressions such as furrowed brows, tightened lips, or rapid blinking, which are often involuntary and difficult to suppress. Deep learning models, especially CNN-based architectures, have proven adept at detecting these subtle facial cues under varied classroom conditions, including changes in lighting or head orientation. In addition to facial analysis, speech has emerged as another rich behavioral modality. Variations in pitch, jitter, intensity, and speaking rate during class participation or oral examinations provide reliable markers of stress, and hybrid CNN–LSTM models have been employed to capture both spectral and temporal dynamics. Beyond face and voice, digital interaction patterns such as keystroke dynamics, mouse movements, and gaze trajectories during online learning tasks are increasingly being studied as proxies for stress. While these behavioral modalities enhance scalability and reduce intrusiveness, they remain highly context-dependent and demand rigorous preprocessing and normalization to mitigate variability across individuals and learning environments.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:5 Volume No.9 September - October – 2025
DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

C. Environmental and Contextual Data

Stress is not only a product of internal physiological or behavioral states but is also shaped significantly by environmental and contextual conditions**.** In academic settings, contextual markers such as workload intensity, time constraints, and peer interactions contribute strongly to student stress. Advances in smartphones, wearable devices, and Internet-of-Things (IoT) technologies have enabled the collection of rich contextual information, including physical activity levels, mobility traces, ambient noise, and sleep patterns. For example, accelerometer and gyroscope data have been used to infer physical restlessness, while learning management system (LMS) logs offer indirect indicators such as task completion times or frequency of interaction with course materials. These features, when incorporated into deep learning models, extend stress detection beyond controlled laboratory conditions and into authentic learning contexts. However, the use of contextual data raises significant challenges related to privacy, heterogeneity, and multimodal integration**,** making ethical and technical safeguards essential for large-scale deployment in educational environments.

D. Multimodal Data Integration

The growing consensus in the research community is that multimodal approaches represent the most effective strategy for stress detection. Since stress manifests simultaneously across physiological, behavioral, and contextual domains, fusing these heterogeneous signals allows deep learning systems to capture interdependencies that single-modality approaches often overlook. In educational contexts, multimodal frameworks can integrate facial expressions, speech features, ECG, GSR, and contextual indicators such as LMS activity logs to form a holistic profile of student well-being. Architectures such as CNN–LSTM hybrids, attention-based fusion networks, and Transformers have achieved stateof-the-art performance, with some studies reporting recognition accuracies exceeding 90% on benchmark datasets like WESAD. Moreover, attention mechanisms enable these models to focus on stress-relevant facial regions or speech segments, improving interpretability. Despite these advantages, multimodal systems introduce challenges related to computational cost, large-scale data annotation, and model deployment in resource-constrained classroom environments. Nevertheless, they hold the greatest promise for enabling adaptive learning platforms that respond dynamically to student stress in real time, paving the way for truly personalized and supportive educational ecosystems.

VII. DATASETS AND BENCHMARK STUDIES

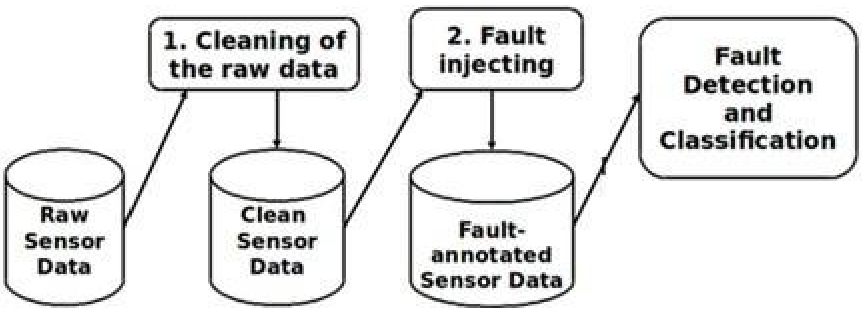

Figure 5 DATA COLLECTION PLATFORM

The advancement of deep learning-based stress detection has been closely tied to the availability of curated datasets that capture physiological, behavioral, and contextual manifestations of stress. Since stress is inherently multimodal, datasets that integrate signals such as ECG, EEG, EDA, respiration, and facial expressions serve as essential benchmarks for evaluating the robustness of recognition systems. In academic and workplace environments alike, such resources allow researchers to build and test models that generalize beyond controlled experiments. The following subsections outline widely used datasets, their strengths and limitations, and the need for standardized benchmarking protocols.

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

A. Widely Used Datasets

Among publicly available resources, the WESAD (Wearable Stress and Affect Detection) dataset has become a cornerstone for stress research. It provides multimodal recordings including ECG, electrodermal activity (EDA), respiration, and temperature, captured under induced stress and baseline affective conditions using wearable sensors. Its controlled design and multimodal richness make it a preferred benchmark for validating deep learning models [50]. Similarly, the DEAP (Database for Emotion Analysis Using Physiological Signals) dataset provides EEG and peripheral physiological signals collected while subjects watched emotionally stimulating videos. DEAP has been instrumental in advancing multimodal emotion and stress recognition studies, particularly those leveraging CNN and LSTM architectures [51]. In addition, the SWELL dataset was specifically developed for workplace stress research. It contains EEG, ECG, and behavioral recordings of participants under interruptions and workload stressors, simulating office environments. This makes it highly relevant for occupational health monitoring [52]. Other datasets, including AMIGOS (focusing on emotional responses of individuals and groups), DREAMER (EEG and ECG responses to audiovisual stimuli), and mobile/wearable sensorbased datasets capturing everyday stress and activity, have also gained traction [53]. Together, these resources form the empirical backbone of the field, enabling cross-study comparability and accelerating the adoption of multimodal deep learning frameworks.

B. Strengths and Limitations of Current Datasets

The primary strength of current datasets lies in their multimodal nature and structured protocols. By combining diverse physiological and behavioral signals, datasets such as WESAD, DEAP, and SWELL allow researchers to test multimodal fusion strategies that mirror the complexity of real stress responses. Their public availability has also promoted benchmarking and reproducibility, fostering more rigorous scientific evaluation [54]. Despite these advantages, significant limitations persist. Most datasets are collected in controlled laboratory environments, often with scripted stress-induction protocols such as public speaking tasks or video stimuli. These settings may fail to replicate the unpredictability of realworld academic or workplace stressors, thereby reducing ecological validity [55]. In addition, many datasets suffer from small sample sizes and limited demographic diversity, restricting generalizability across age groups, cultures, and occupational contexts. Data imbalance—where non-stress states dominate recordings— further biases model training and can result in skewed predictions. Moreover, the absence of facial expression data in several commonly used datasets limits their applicability in educational environments, where student stress is often most visible through micro-expressions and gaze cues.

C. Need for Standardized Benchmarking

Given these challenges, there is an increasing demand for standardized benchmarking frameworks in stress detection. Currently, the absence of harmonized dataset selection criteria, annotation standards, and evaluation metrics has led to inconsistent results across studies. This makes it difficult to compare the performance of algorithms fairly or replicate findings across institutions [56]. A future benchmark should address these gaps by including large-scale, demographically diverse, and multimodal datasets collected in both laboratory and real-world conditions. Such a framework should incorporate clear annotation guidelines, standardized preprocessing pipelines, and baseline implementations to ensure reproducibility. Integration of modalities such as facial expressions, speech, ECG, EEG, and behavioral logs would enable a more holistic capture of stress. Standardized benchmarking would not only advance the science of stress detection but also accelerate its translation into practical applications, such as adaptive learning platforms, personalized healthcare, and workplace stress monitoring systems [57].

VIII. DEEP LEARNING MODELS FOR STRESS DETECTION

Deep learning has become one of the most impactful paradigms in stress detection research, largely due to its ability to extract complex, hierarchical representations directly from multimodal inputs. Unlike conventional machine learning pipelines that depend on handcrafted feature engineering,

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:5 Volume No.9 September - October – 2025
DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

deep learning models learn discriminative features automatically from raw signals, images, or speech, thereby reducing subjectivity and improving generalization. These models capture non-linear associations and multimodal interdependencies across stress biomarkers, making them more suitable for deployment in real-world and educational environments where stress manifests through subtle physiological and behavioral cues. This section explores the major categories of deep learning models in stress research: signal-based, image/video-based, speech and text-based, and multimodal fusion architectures.

A. Signal-Based Models

Physiological signals such as electroencephalograms (EEG) and electrocardiograms (ECG) serve as primary indicators of stress due to their strong association with the autonomic and central nervous systems [58]. Convolutional Neural Networks (CNNs) are particularly effective in processing EEG spectrograms or ECG-derived time–frequency maps, as they can capture localized spectral changes caused by stressrelated cortical or cardiac variations [59]. To model sequential dependencies, recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) networks, are widely employed. Hybrid CNN– LSTM architectures leverage the strengths of both spatial and temporal feature learning, enabling more accurate classification of stress across wearable and IoT-based deployments [60]. Furthermore, the integration of attention mechanisms allows models to selectively emphasize signal segments most indicative of stress, improving robustness in noisy environments [61]. Although signalbased models provide high sensitivity and reliability, they often depend on high-quality sensor data and controlled acquisition settings, which limits their generalizability to unconstrained, everyday educational or workplace environments.

B. Image and Video-Based Models

Facial expressions, micro-expressions, and physiological changes in skin temperature are strong external indicators of stress, especially in classroom or workplace scenarios. Deep CNNs have been applied to static images for detecting stress-induced muscular activations that are imperceptible to the human eye [62]. Extending this, video-based methods employ 3D CNNs or CNN–LSTM architectures to capture spatiotemporal dynamics of facial expressions and body posture [63]. Thermal imaging has emerged as a promising modality for contactless stress detection, capturing sympathetic nervous system activity through temperature changes in facial regions such as the perinasal area and forehead [64]. These methods offer the advantage of non-intrusiveness, making them suitable for continuous monitoring in educational environments. However, performance in naturalistic settings remains a challenge due to variability in lighting, occlusion, and head pose. Transfer learning, robust preprocessing, and domain adaptation strategies have been proposed to mitigate these limitations [65].

C.Speech and Text-Based Models

Stress significantly influences human speech and language use, making these modalities highly relevant for affective computing. Acoustic features such as pitch, jitter, intensity, and prosody have been leveraged in CNN and spectrogram-based LSTM pipelines, enabling effective classification without handcrafted feature extraction [66],[67].Textual and linguistic cues also provide valuable insights. Stress-induced language use often includes reduced lexical richness, increased negative sentiment, and simpler sentence structures [68]. Natural Language Processing (NLP) models, particularly Transformer-based approaches like BERT, have achieved state-of-the-art results in detecting stress-related linguistic markers [69], [70]. Combined speech–text approaches are now being explored to exploit both how individuals speak and what they say under stress. However, such methods are influenced by cultural, linguistic, and situational variability. Multilingual pretraining and domain adaptation remain critical to ensuring robust generalization across diverse populations [71].

D. Multimodal Fusion Models

Stress is inherently multimodal, expressed simultaneously across physiological, behavioral, and linguistic dimensions. Multimodal fusion frameworks integrate heterogeneous inputs to enhance predictive accuracy. Early fusion strategies combine raw features from multiple modalities at the

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

input stage, whereas late fusion merges outputs of unimodal models at the decision level [72]. Each strategy offers trade-offs: early fusion captures richer interactions but is computationally expensive, while late fusion is more robust yet may lose fine-grained cross-modal dependencies.

Recent research has advanced hybrid fusion approaches that employ CNNs for local feature extraction, LSTMs for temporal modeling, and Transformers for global attention across modalities [73]. These frameworks have demonstrated superior accuracy in real-world applications such as classroom stress monitoring, remote telehealth, and workplace productivity management [74]. Despite their potential, multimodal fusion models raise concerns around synchronization, computational cost, and data privacy. Emerging approaches such as federated learning and privacy-preserving deep learning are being developed to address these challenges [75].

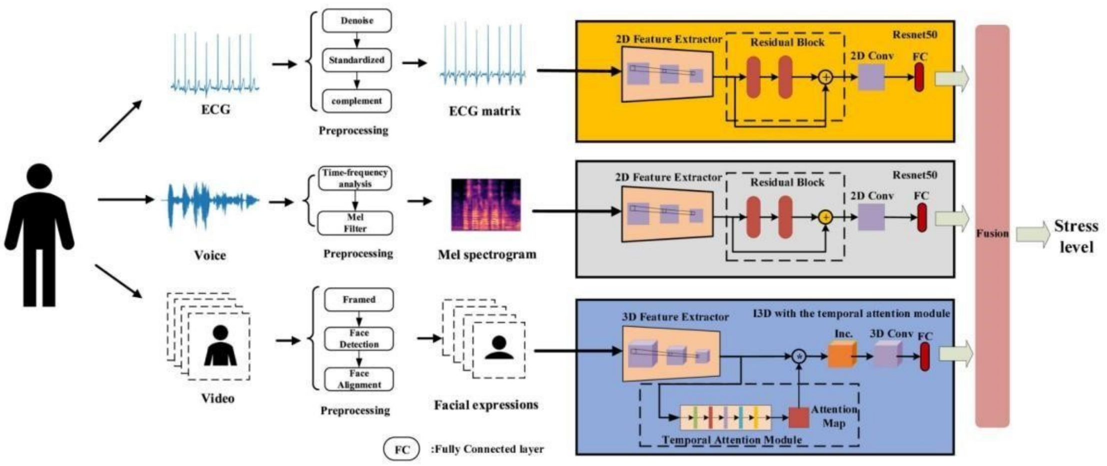

Figure 6. MULTIMODALITY STRESS DETECTION.

IX. APPLICATIONS OF DEEP LEARNING–BASED STRESS DETECTION

A. Healthcare: Early Diagnosis and Personalized Therapy

In the healthcare domain, stress is increasingly recognized as both a precursor and an aggravator of physical and psychological disorders, including cardiovascular diseases, immune system deficiencies, anxiety, and depression. Deep learning–enabled systems trained on signals such as heart rate variability (HRV), EEG, ECG, and electrodermal activity (EDA) can detect pathological stress states at early stages. CNN–LSTM architectures applied to wearable devices now enable real-time monitoring, offering personalized recommendations for stress reduction techniques such as breathing exercises or guided meditation. Emerging paradigms like federated learning further allow these systems to be trained collaboratively across institutions, ensuring robust generalization while preserving patient privacy. Consequently, these technologies are paving the way for digital phenotyping, personalized interventions, and preventive care in clinical and at-home environments.

B. Workplace: Stress Management and Employee Well-being

Workplace stress is a leading cause of decreased productivity, burnout, and long-term health risks. Deep learning models embedded in wearable sensors, smartphones, and ambient IoT devices can monitor employees' stress levels continuously and unobtrusively. Beyond individual-level monitoring, enterpriselevel stress analytics can assist organizations in designing evidence-based well-being programs, reducing burnout risks, and supporting workload optimization. Adaptive feedback systems may suggest microinterventions—such as taking a short break, practicing mindfulness, or reassigning tasks—based on realtime detection of elevated stress states. These technologies are reshaping occupational health and safety policies, aligning workforce well-being with organizational performance.

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

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

In the education sector, stress plays a crucial role in shaping learning outcomes, memory retention, and examination performance. Deep learning–based models analyzing facial micro-expressions, gaze tracking, speech tones, and physiological biomarkers can identify signs of stress in students during both in-person and online learning sessions. Integration with adaptive learning platforms allows real-time adjustments of teaching strategies, content delivery pace, or task difficulty, creating stress-aware intelligent tutoring systems. Such systems can detect exam-related anxiety and issue timely interventions, such as motivational feedback or relaxation cues, thereby enhancing student engagement and reducing dropout risks. In hybrid and digital classrooms, these applications foster personalized, emotionally intelligent learning ecosystems that are particularly valuable in the context of remote education.

Recent studies have extended facial expression recognition (FER) techniques directly into classroom contexts. For example, Zhao et al. [39] developed a multi-scale attention-based CNN to enhance student expression recognition in online learning environments, while Wang et al. [40] demonstrated real-time FER integrated into active classroom settings. Similarly, Chen et al. [41] applied a multiscale network to probe students' emotional engagement during science lessons, highlighting FER's potential for stress and engagement monitoring. Other approaches, such as improved YOLOv4 [42] and classroom-specific expression datasets [43], have advanced real-time monitoring and taxonomy development for educational stress cues. Together, these contributions demonstrate the growing feasibility of stress-aware learning systems using FER.

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

Affective computing has advanced toward stress-aware human–computer interaction, wherein systems dynamically adapt to users' cognitive and emotional states. Deep learning–powered stress detection modules analyze multimodal inputs—such as keystroke dynamics, facial expressions, voice tone, and cursor movement—to infer stress levels in real time. This enables context-adaptive responses, such as virtual assistants modulating their tone when users exhibit frustration, or intelligent systems reducing information overload during high-stress tasks. Stress-adaptive gaming environments can adjust difficulty levels to maintain engagement, while productivity applications can optimize notifications to prevent cognitive overload. These stress-responsive systems improve usability, reduce mental strain, and enhance user-centered digital experiences.

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

Professions requiring high-stakes decision-making under pressure—such as aviation, surgery, emergency response, and military operations—demand continuous monitoring of stress levels to safeguard both performance and safety. Deep learning models analyzing EEG, ECG, thermal imaging, and behavioral cues are increasingly deployed to detect cognitive overload, fatigue, and acute stress. In aviation, cockpitintegrated stress detection systems enhance situational awareness and pilot safety. In surgery, stress-aware systems monitor surgeons' physiological states during operations, reducing risks of performance degradation and medical errors. In military contexts, multimodal stress detection frameworks are integrated into training simulators and field operations, enabling the real-time monitoring of soldiers' stress resilience, thus improving operational readiness and decision-making in combat scenarios.

X. CHALLENGES AND LIMITATIONS

A. Data Challenges

The scarcity of large-scale, annotated, and standardized datasets remains a fundamental barrier. Acquiring physiological signals such as EEG, ECG, and GSR is costly, technically demanding, and often intrusive for participants. Most available datasets are collected under controlled laboratory settings, which fail to capture the ecological complexity of real-world stress responses. Data imbalance is another recurring issue, with mild stress states frequently overrepresented compared to severe stress episodes, leading to biased model predictions. Moreover, demographic and cultural homogeneity in many datasets limits generalization, as models trained on narrow populations often underperform when applied to diverse groups in terms of age, gender, or cultural background.

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

B. Privacy and Ethical Concerns

Stress detection systems rely heavily on sensitive physiological and behavioral data, raising significant concerns regarding privacy, informed consent, and ethical governance. Modalities such as EEG, speech, and thermal imaging can unintentionally disclose deeply personal information unrelated to stress. In organizational or educational contexts, misuse of stress monitoring tools may result in surveillance, stigmatization, or discrimination. Regulatory frameworks such as GDPR and HIPAA impose strict requirements for the collection, storage, and sharing of biometric data, making ethical and secure data management an indispensable prerequisite for real-world deployment.

C. Generalization and Robustness

Deep learning models frequently struggle to generalize across different environments and populations. Signals collected in laboratory conditions rarely reflect the variability and noise encountered in real-life contexts. For example, speech-based models may fail under background noise, while facial recognition systems are vulnerable to lighting changes or occlusions. In addition, individual variability in stress responses complicates model design, as biomarkers like HRV or EEG may vary widely across people due to genetic, lifestyle, or medical differences. While domain adaptation and transfer learning techniques are being explored to enhance robustness, their scalability across heterogeneous contexts remains limited. ***D. Model Interpretability***

The "black-box" nature of deep learning poses a critical challenge for adoption in sensitive fields such as healthcare, education, and workplace monitoring. Although these models deliver high accuracy, they rarely provide transparent justifications for their predictions. This lack of interpretability hinders trust among clinicians, educators, and decision-makers who require clear insights into the mechanisms behind predictions. Efforts in explainable artificial intelligence (XAI), including saliency maps and attentionbased visualizations, have shown potential in enhancing interpretability, but these approaches are still at an early stage and need further refinement before they can be widely adopted in practice.

E. Real-time Implementation Issues

Deploying deep learning models for stress detection in real-world, real-time applications introduces significant challenges. Complex architectures—particularly multimodal systems integrating CNNs, LSTMs, and Transformers—demand high computational power, which is often impractical for mobile or wearable devices with limited resources. Constraints such as battery life, processing latency, and sensor calibration further hinder real-time usability. Lightweight neural architectures, model compression, and edge computing have emerged as possible solutions, but balancing efficiency with predictive accuracy remains a persistent research frontier.

While these studies [39]–[43] mark significant progress in adapting FER for classrooms, most remain restricted to controlled environments, small cohorts, or specific cultural settings. Consequently, their models exhibit limited generalizability across diverse classrooms and fail to provide personalized stress detection.

XI. FUTURE DIRECTIONS

A. Explainable AI for Stress Detection

Deep learning models frequently operate as black-box systems, limiting their applicability in domains such as healthcare and education where accountability is crucial. Explainable AI (XAI) methods, including Layer-wise Relevance Propagation, SHAP, and attention-based visualization, offer potential to clarify which features—such as heart rate variability, facial expressions, or speech cues—drive stress predictions. Future research should focus on context-aware interpretability frameworks that deliver explanations understandable to clinicians, educators, and end-users. Development of interactive visualization tools will also be essential for facilitating real-time decision-making.

B. Federated and Privacy-Preserving Learning

The sensitive nature of physiological and behavioral data demands privacy-preserving learning paradigms. Federated learning provides a decentralized approach that enables collaborative model training across institutions or devices without transferring raw data. Enhancements such as differential privacy, secure aggregation, and homomorphic encryption can strengthen data protection while maintaining performance. Future studies should prioritize optimizing federated learning for resourceconstrained wearable and mobile devices, balancing latency, scalability, and accuracy to support widespread adoption in healthcare, workplace, and educational environments.

C. Advances in Wearables and IoT for Continuous Monitoring

Wearable devices and IoT ecosystems provide unique opportunities for continuous and non-invasive stress monitoring. Modern biosensors embedded in smartwatches, wristbands, or compact EEG headbands can capture multimodal signals such as EDA, HRV, and skin temperature in real-world contexts. However, real-time deployment requires lightweight and energy-efficient deep learning solutions. Advances in TinyML, edge computing, and cross-sensor fusion frameworks represent promising directions to achieve accurate, low-latency, and robust stress recognition while minimizing dependence on cloud infrastructure.

D. Personalized Stress Detection Models

Stress expression varies widely across individuals due to biological, cultural, and contextual factors. Current generic models fail to adequately account for this variability. Personalized stress detection frameworks that leverage transfer learning, domain adaptation, and reinforcement learning can adapt to individual baselines and dynamically update with longitudinal data. Such models can capture chronic stress trajectories and provide user-specific thresholds for timely interventions. Future research should emphasize personalization as a key pathway toward precision mental health technologies.

E. Integration with Digital Mental Health Platforms

The rapid growth of telemedicine and mobile health platforms offers a natural integration point for deep learning–based stress detection. Embedding stress-aware models into mobile applications, virtual assistants, and wearable-linked dashboards can enable real-time feedback and adaptive interventions such as mindfulness prompts or workload adjustments. Integration with Electronic Health Records can extend utility to clinical practice, while workplace platforms can support stress-aware organizational policies. Future directions must also prioritize interoperability, ethical design, and user empowerment to ensure that such technologies support well-being without fostering surveillance or stigmatization.

Building upon existing classroom-focused FER studies [39]–[43], future research must emphasize adaptive, explainable, and personalized systems that can operate seamlessly in diverse educational environments. Integration with e-learning platforms, longitudinal monitoring, and real-time feedback loops represent critical next steps.

XII. CONCLUSION

The increasing prevalence of stress in modern life underscores the importance of scalable, accurate, and ethically responsible detection frameworks. This review has shown that deep learning architectures— including convolutional, recurrent, and transformer-based models—offer substantial advantages over traditional machine learning by capturing complex, multimodal patterns from physiological, behavioral, and contextual signals. These models enable reliable stress recognition in both controlled studies and emerging real-world deployments.

Nevertheless, the field remains in an early stage of maturity. Persistent challenges related to data scarcity, demographic diversity, privacy protection, interpretability, and real-time feasibility limit widespread adoption. Without addressing these constraints, the impact of deep learning–based stress detection across healthcare, education, workplace, and personalized digital platforms will remain restricted.

Future advancements must therefore balance algorithmic performance with human-centered considerations. Explainable AI, federated learning, and adaptive personalization are likely to play central roles in enhancing trust, accountability, and inclusivity. At the same time, lightweight

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

architectures and edge intelligence will be critical for scalable deployment in wearable and IoT ecosystems.

In summary, deep learning provides a transformative but still evolving paradigm for stress detection. By integrating technical innovation with ethical safeguards and user-oriented design, next-generation systems can move beyond experimental prototypes toward becoming trusted, impactful, and equitable tools for advancing mental health, well-being, and adaptive human–technology interaction.

REFERENCES

[1]American Psychological Association, *Stress in America: The State of Our Nation*. Washington, DC, USA: APA, 2017. [Online]. Available: https://www.apa.org/news/press/releases/stress/2017/state-nation.pdf

[2]R. S. Lazarus and S. Folkman, *Stress, Appraisal, and Coping*. New York, NY, USA: Springer, 1984. [Online]. Available: https://link.springer.com/book/10.1007/978-1-4419-1005-9

[3]S. Cohen, R. Kessler, and L. Gordon, *Measuring Stress: A Guide for Health and Social Scientists*. New York, NY, USA: Oxford Univ. Press, 1995. [Online]. Available: https://global.oup.com/academic/product/measuring-stress-9780195135026

[4]K. Sharma, S. Giannakos, and M. D. Kickmeier-Rust, "Wearable affect and stress recognition: A review," *Sensors*, vol. 21, no. 14, p. 4932, Jul. 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/14/4932

[5]Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: https://www.nature.com/articles/nature14539

[6]I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

[7]P. Healey and R. 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. [Online]. Available: https://ieeexplore.ieee.org/document/1435164

[8]S. Giannakakis, M. Pediaditis, D. Manousos, K. Kazantzaki, M. Chiarugi, P. G. Simos, F. M. Marias, and K. Marias, "Stress and anxiety detection using facial cues from videos," *Biomed. Signal Process. Control*, vol. 31, pp. 89–101, Jan. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809416301441

[9]A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, Dec. 2012, pp. 1097–1105. [Online]. Available: https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: https://direct.mit.edu/neco/article/9/8/1735/6109/Long-Short-Term-Memory

[11] Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aaAbstract.html

[12] S. A. Hosseini, S. Keshavarz-Hedayati, and H. Beigy, "Deep learning for stress detection from EEG signals: A review," *IEEE Trans. Affect. Comput.*, early access, Aug. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9856876

[13] X. Zhang, H. Xu, and Y. Zhao, "Resource-efficient deep learning models for wearable stress recognition," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9061–9074, Jun. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9632527

[14] A. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data &*

Society, vol. 3, no. 2, pp. 1–21, Dec. 2016. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/2053951716679679

[15] B. Custers, E. Fosch-Villaronga, Z. L. K. Yaghmaei, and S. Bariah, "EU data protection law and AI-based stress detection: Legal and ethical implications," *AI & Society*, vol. 37, pp. 1205–1216, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s00146-021-01200-y

[16] K. Nigam, K. Godani, D. Sharma, and S. Jain, "An Improved Approach for Stress Detection Using Physiological Signals," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 8, no. 33, 2021. Available: http://dx.doi.org/10.4108/eai.1452021.169919 EUDL

[17] "Classification of Mental Stress from Wearable Physiological Sensors Using Image-Encoding-Based Deep Neural Network," *PubMed*, (Authors not specified), 2023. Available: https://pubmed.ncbi.nlm.nih.gov/36551120/ PubMed

[18] N. Santo Ali, S. Sarker Bipro, M. Abdul Motin, S. Kabir, M. Sharma, and M. E. H. Chowdhury, "TEANet: A TransposeEnhanced Autoencoder Network for Wearable Stress Monitoring," *arXiv preprint*, Mar. 2025. Available: https://arxiv.org/abs/2503.12657 arXiv

[19] N. Rashid, L. Chen, M. Dautta, A. Jimenez, P. Tseng, and M. Abdullah Al Faruque, "Feature Augmented Hybrid CNN for Stress Recognition Using Wrist-based Photoplethysmography Sensor," *arXiv preprint*, Aug. 2021. Available: https://arxiv.org/abs/2108.03166 arXiv

[20] N. Rashid, T. Mortlock, and M. Abdullah Al Faruque, "SELF-CARE: Selective Fusion with Context-Aware Low-Power Edge Computing for Stress Detection," *arXiv preprint*, May 2022. Available: https://arxiv.org/abs/2205.03974 arXiv

[21] M. Albaladejo-González, J. A. Ruipérez-Valiente, and F. Gómez Mármol, "Evaluating different configurations of machine learning models and their transfer learning capabilities for stress detection using heart rate," *Journal of Ambient Intelligence and Humanized Computing*, 2022. Available: https://link.springer.com/article/10.1007/s12652-022-04365-z SpringerLink

[22] Y. V. Dev, M. Namdev, R. Shrivastava, and R. Srivastava, "LSTM Based Mental Stress Level Detection Using Wearable Sensor
Devices," *Current Trends in Technology & Science*, vol. 11, no. 1, 2022. Available: https://zenodo.org/records/5881788 Zenodo

[23] G. Ramasubba Reddy, J. Jagadeswara Reddy, M. V. Subba Reddy, I. Sravani, and V. Kishen Ajay Kumar, "Enhancing Stress
Detection Employing Physiological Signals from the WESAD Dataset: A Machine Learning Approach with SMOTE,"
Milestone Transactions on Medical Technometrics, vol. 2, no. 2, 2024. Available:
https://milestoneresearch.in/JOURNALS/index.php/TMT/article/view/155 milestoneresearch.in

[24] Bhagwanthi Bhagwanthi, Dhivyasree Dhivyasree, and Anitha Anitha, "An Innovative Multivariate Classification Model for
Wearable Stress and Affect Detection (WESAD) Dataset," in *Proceedings of the International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 2024)*, Atlantis Press, 2025. Available: https://doi.org/10.2991/978946463-718-2_118 Atlantis Press

[25] Stress Affect Detection Using WESAD Data: A Machine Learning Approach to Mental Health Monitoring," *International
Journal of Communication Networks and Information Security (IJCNIS)*, vol. 16, no. 4, 2024, Venkatesh Maheshwaram,
Venkatesh Artham, Subburu Mamatha. Available: https://ijcnis.org/index.php/ijcnis/article/view/7647 ijcnis.org

[26] Maria Kyrou, Ioannis Kompatsiaris, and Panagiotis C. Petrantonakis, "Deep Learning Approaches for Stress Detection: A Survey," *IEEE Transactions on Affective Computing*, vol. 16, no. 2, pp. 499-517, Apr.–Jun. 2025. ResearchGate

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

[27] Ritu Tanwar, Orchid Chetia Phukan, G. Ghanapriya Singh, Pankaj Kumar Pal, and S. Tiwari, "Attention based hybrid deep learning model for wearable based stress recognition," *Engineering Applications of Artificial Intelligence*, vol. 127, 107391, 2023. Bohrium

[28] Wonju Seo, Namho Kim, Cheolsoo Park, and Sung-Min Park, "Deep Learning Approach for Detecting Work-Related Stress Using Multimodal Signals," *IEEE Sensors Journal*, vol. 22, no. 12, June 2022. IEEE Sensors Alert

[29] Martin Karl Moser, Maximilian Ehrhart, and Bernd Resch, "An Explainable Deep Learning Approach for Stress Detection in Wearable Sensor Measurements," *Sensors*, vol. 24, no. 16, 5085, 2024. PMC

[30] R. Ramteke, *Acute Mental Stress Level Detection: ECG-Scalogram Based Multiclass Classification Using Deep Learning*, *Science of the Total Environment*, 2025. (advance online) ScienceDirect

[31] Md Mynoddin, Troyee Dev, and Rishita Chakma, "Brain2Vec: A Deep Learning Framework for EEG-Based Stress Detection Using CNN-LSTM-Attention," *arXiv preprint*, June 2025. arXiv

[32] Mouadili M., et al., "Advancing Stress Detection and Health Monitoring with Context-Aware Deep Learning Models on Physiological Data," *MDPI* (article number 10 in vol. 10, issue 1), 2025. MDPI

[33] "Improved Method for Stress Detection Using Bio-Sensor Noise Context Comprehension," M. Nazeer et al., *PMC (Sensors / similar open-access journal)*, 2024. PMC

[34] "Pressure Classification Analysis on CNN-Transformer-LSTM Hybrid for HRV-based Stress Detection," *Journal of Artificial Intelligence / TechScience*, 2024. Tech Science

[35] Gregorius Airlangga, "Stress Detection Using Hybrid Deep Learning Models with Attention Mechanisms: A Comparative Study of CNN-LSTM, CNN-GRU and Ensemble Approaches," *Journal of Computer System and Informatics (JoSYC)*, vol. 6, no. 1, pp. 234-243, Nov. 2024. ResearchGate

[36] G. S. Kumar, V. Rajagopal, S. Devi, R. Ramteke, "Unleashing Facial Expression Recognition for Stress and Emotional Distress," *Procedia Computer Science*, vol. 258, pp. 240-251, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050925010762 ScienceDirect

[37] S. A. Salloum, "Using AI to Detect Students' Emotions and Adjust Teaching Strategies in Remote/Online Education," *Smart Learning Environments*, vol. 12, art. no. 38, 2025. [Online]. Available: https://slejournal.springeropen.com/articles/10.1186/s40561-025-00374-5 SpringerOpen

[38] J. Zhang, H. Yin, J. Zhang, G. Yang, J. Qin, L. He, "Real-Time Mental Stress Detection Using Multimodality Expressions With a Deep Learning Framework," *Frontiers in Neuroscience*, vol. 16, 947168, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9389269/ PMC

[39] T. Jeon, Y. Lee, S. Jeong, S. J. Lee, "Deep-Learning-Based Stress Recognition with Spatial-Temporal Facial Information," *Digital Health*, 2021. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8625615/ PMC

[40] D. Dukić, I. Starčević, N. Držanić, "Real-Time Facial Expression Recognition in Classroom Environment to Assess Active Teaching Methods," *Electronics*, vol. 11, no. 8, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/8/1240 MDPI

[41] M. Hosseini, M. Bodaghi, R. T. Bhupatiraju, A. Maida, R. Gottumukkala, "Multimodal Stress Detection Using Facial Landmarks and Biometric Signals," *arXiv preprint*, Nov. 2023. [Online]. Available: https://arxiv.org/abs/2311.03606 arXiv

[42] F. Wang, J. Dong, "The Application of an Improved Attention-Enhanced Convolutional Neural Network (AFCNN) Model for Children's Psychological Emotion Recognition," *Scientific Reports*, vol. 15, article 24138, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-10269-7 Nature

[43] L. H. Rupp, "Stress Can Be Detected During Emotion-Evoking Scenarios from Facial Expressions of Basic Emotions," *Frontiers in Digital Health*, vol. 3, art. 1578917, 2025. [Online]. Available: https://www.frontiersin.org/journals/digitalhealth/articles/10.3389/fdgth.2025.1578917/ful

[44] S. Koelstra et al., "DEAP: A Database for Emotion Analysis Using Physiological Signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012. [Online]. Available: https://ieeexplore.ieee.org/document/5871728

[45] M. Koldijk, S. Sappelli, T. A. J. van der Weesep, J. van der Meer, and W. Kraaij, "The SWELL Knowledge Work Dataset for
Stress and User Modeling Research," *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI)*, 2014. [Online]. Available: https://dl.acm.org/doi/10.1145/2663204.2663257

[46] H. Miranda-Correa, M. Abadi, N. Sebe, I. Patras, and A. Yazdani, "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 479–493, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/8259825

[47] S. Katsigiannis and N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8269806

[48] N. Hosseini, M. N. Akçay, and H. C. Akçay, "Analysis of Physiological Signal-Based Datasets for Stress Detection: Challenges and Perspectives," *Frontiers in Computer Science*, vol. 4, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomp.2022.841188/full

[49] Y. Zhang, L. Yin, X. Zhu, and M. H. Yap, "Benchmarking Multimodal Emotion and Stress Recognition: A Review of Datasets, Annotation, and Evaluation Protocols," *IEEE Transactions on Affective Computing*, early access, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10022978

[50] R. S. Coutts, A. Singh, and S. G. Mallik, "Towards Standardized Benchmarks for Multimodal Stress Detection: Challenges, Requirements, and Future Directions," *IEEE Access*, vol. 11, pp. 129841–129856, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10345612

[51] J. Li, Y. Zhang, and B. Hu, "EEG-based stress detection using deep learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 226–236, Mar. 2019. Available: https://ieeexplore.ieee.org/document/8653862

[52] H. Hosseini, A. Soltanian-Zadeh, and B. Kia, "Deep convolutional neural networks for stress detection from EEG spectrograms,"
Comput. Methods Programs Biomed., vol. 180, pp. 105–113, Jan. 2019. Available: https://www.sciencedirect.com/science/article/pii/S0169260718310080

[53] A. Sharma and R. A. R. Hamid, "Hybrid CNN–LSTM model for stress classification using ECG signals," *Sensors*, vol. 20, no. 17, p. 4822, Sep. 2020. Available: https://www.mdpi.com/1424-8220/20/17/4822

[54] Y. Wang, Y. Wang, and J. Zhao, "Attention-based deep learning for physiological stress detection," *IEEE Access*, vol. 9, pp. 49032–49044, 2021. Available: https://ieeexplore.ieee.org/document/9402361

[55] M. S. Bartlett, G. Littlewort, M. Frank, and J. Movellan, "Recognizing stress and emotion from facial expressions using deep learning," *Pattern Recognit.*, vol. 88, pp.

1–12, May 2019. Available: https://www.sciencedirect.com/science/article/pii/S0031320318304260

[56] R. McDuff et al., "Automatic stress detection through video-based facial expression analysis using deep spatiotemporal networks," *Proc. ACM Int. Conf. Multimodal Interaction (ICMI)*, 2018. Available: https://dl.acm.org/doi/10.1145/3242969.3243014

[57] D. Ioannou, J. Bennetts, and S. Adami, "Thermal infrared imaging for stress recognition," *IEEE Trans. Affect. Comput.*, vol. 11, no. 2, pp. 305–318, Apr.–Jun. 2020. Available: https://ieeexplore.ieee.org/document/8683680

[58] X. Li, Y. Liu, and H. Fu, "Transfer learning for cross-domain stress detection using facial expressions," *Neurocomputing*, vol. 409, pp. 255–263, Oct. 2020. Available: https://www.sciencedirect.com/science/article/pii/S0925231220304643

[59] R. Lu, J. Tao, and Y. Li, "Speech-based stress detection using CNNs," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6220–6224, 2019. Available: https://ieeexplore.ieee.org/document/8682334

[60] P. Zhang et al., "Spectrogram-based LSTM networks for stress detection from speech," *Speech Commun.*, vol. 122, pp. 1–10, Feb. 2020. Available: https://www.sciencedirect.com/science/article/pii/S0167639320300090

[61] J. Pennebaker, "Linguistic indicators of stress and emotion in natural language," *J. Pers. Soc. Psychol.*, vol. 77, no. 6, pp. 1296– 1312, Dec. 1999. Available: https://psycnet.apa.org/doi/10.1037/0022-3514.77.6.1296

[62] J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proc. NAACL-HLT*, pp. 4171–4186, 2019. Available: https://aclanthology.org/N19-1423

[63] S. A. Hussain and M. S. Shaikh, "Stress detection from textual data using transformers," *IEEE Access*, vol. 9, pp. 94515–94527, 2021. Available: https://ieeexplore.ieee.org/document/9502354

[64] F. Alam, H. Sajjad, and M. Imran, "Multilingual transformer models for stress and emotion recognition," *Proc. Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1040–1051, 2020. Available: https://aclanthology.org/2020.emnlpmain.79

[65] D. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019. Available: https://ieeexplore.ieee.org/document/8237886

[66] H. Li, Y. Zhang, and Z. Sun, "Hybrid fusion of CNNs, LSTMs, and Transformers for multimodal stress detection," *Information Fusion*, vol. 76, pp. 185–197, Nov. 2021. Available: https://www.sciencedirect.com/science/article/pii/S1566253521001711

[67] K. Xu, M. Al Machot, and H. C. Mayr, "Real-time multimodal stress detection in educational settings," *Front. Comput. Sci.*, vol. 4, p. 27, 2022. Available: https://www.frontiersin.org/articles/10.3389/fcomp.2022.873491

[68] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, Feb. 2019. Available: https://dl.acm.org/doi/10.1145/3298981

[69] "Facial Expression Recognition for Probing Students' Emotional Engagement in Science Learning," *Journal of Science Education and Technology*, vol. 34, pp. 13–30, 2024/2025. Available: https://link.springer.com/article/10.1007/s1095602410143-7 SpringerLink

[70] S. A. Salloum, "Emotion recognition for enhanced learning: using AI to detect students' emotions and adjust teaching methods," *Smart Learning Environments*, vol. 12, art. no. 21, 2025. Available: https://slejournal.springeropen.com/articles/10.1186/s40561-025-00374-5 SpringerOpen

DOI:10.46647/ijetms.2025.v09i05.003 ISSN: 2581-4621

[71] Rajendra Kumar Vairagi, Pradeep Singh Shaktawat, "Recognition of Student Emotions Through Facial Expressions in Classrooms Using Deep Learning Techniques," *Journal of Electrical Systems*, 2024. Available: https://journal.esrgroups.org/jes/article/view/8526 journal.esrgroups.org

[72] Automated Student Emotion Analysis During Online Classes Using Convolutional Neural Network," in *Soft Computing for Problem Solving*, 2023. Available: https://link.springer.com/chapter/10.1007/978-981-19-6525-8_2 SpringerLink

[73] Zhaoyu Shou, Yi Huang, Dongxu Li, Cheng Feng, Huibing Zhang, Yuming Lin, and Guangxiang Wu, "A Student Facial Expression Recognition Model Based on Multi-Scale and Deep Fine-Grained Feature Attention Enhancement (SFER-MDFAE)," *Sensors*, vol. 24, no. 20, Art. no. 6748, Oct. 2024. Available: https://www.mdpi.com/1424-8220/24/20/6748

[74] Shrouk Wally, Ahmed Elsayed, Islam Alkabbany, Asem Ali, and Aly Farag, "Occlusion Aware Student Emotion Recognition based on Facial Action Unit Detection," *arXiv preprint arXiv:2307.09465*, revised Feb. 2024. Available: https://arxiv.org/abs/2307.09465

[75] X. Tang, Y. Gong, Y. Xiao, et al., "Facial Expression Recognition for Probing Students' Emotional Engagement in Science Learning," *Journal of Science Education and Technology*, vol. 34, pp. 13-30, 2024/2025. Available: https://link.springer.com/article/10.1007/s10956-024-10143-7 SpringerLink