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.2025.74110

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

Deep Learning for Emotion Detection in Classroom Environment

Kavitha M[1], Dr. Jothish Chembath[2]
[1]*Research Scholar,* [2]*Associate Professor, School of Information Science, Presidency University,Bengaluru-560064, Karnataka*

Abstract: Stress has emerged as a critical factor influencing student academic performance, classroom participation, and overall emotional well-being. Early identification of stress is essential for implementing adaptive learning strategies and ensuring timely intervention to improve both mental health and educational outcomes. Conventional assessment approaches, such as self-report questionnaires, observational analysis, and psychological surveys, often lack scalability, objectivity, and the ability to capture real-time variations in student behaviour. Recent advances in artificial intelligence, particularly deep learning, have introduced powerful tools capable of processing multimodal data—including physiological signals, facial expressions, and speech features—to achieve more accurate and dynamic stress detection. This review provides an in-depth analysis of deep learning models applied to classroom stress detection. The discussion covers theoretical models of stress, relevant biomarkers, traditional assessment methods, and modern architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) models, and Transformer frameworks. Furthermore, multimodal fusion strategies, performance evaluation techniques, and real-time classroom applications are examined. The paper also highlights challenges related to data privacy, scalability, and ethical considerations, while outlining future directions with emphasis on explainable AI and integration into smart classroom systems.
Keywords: Stress detection, deep learning, multimodal data, classroom environments, CNN, RNN, LSTM, transformer models, explainable AI.

I. INTRODUCTION

Stress is an increasingly common phenomenon in academic settings and has been closely associated with student performance, motivation, and classroom engagement [1]. In educational environments, unmanaged stress can impair memory, concentration, and decision-making ability, thereby reducing the quality of learning [2]. Addressing this issue requires timely and accurate detection of stress, enabling educators to adopt adaptive teaching methods and provide psychological support.

Traditional approaches to stress detection, such as psychological questionnaires and behavioral observations, have been widely used but are limited by their subjectivity and inability to capture real-time variations [3]. Physiological signal–based methods, including heart rate variability (HRV), galvanic skin response (GSR), and electroencephalography (EEG), offer greater objectivity but suffer from challenges of intrusiveness and scalability [4]. The emergence of deep learning provides a promising solution, as it enables the analysis of complex multimodal data and allows for robust pattern recognition of stress-related indicators [5]. Models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid CNN–LSTM architectures have demonstrated effectiveness in identifying subtle cues from biosignals, speech, and facial expressions [6], [7]. More recently, transformer-based models and multimodal fusion strategies have further enhanced accuracy and reliability, paving the way for deployment in real-time classroom environments [8], [9].

This review article explores the theoretical foundations, data modalities, and deep learning methodologies used for stress detection. It also examines evaluation metrics, real-world applications in classroom settings, and existing challenges. The paper concludes by identifying future research directions, including privacy-preserving approaches, explainable AI, and integration with adaptive learning systems.

II. THEORETICAL BACKGROUND

A. Stress Models

The conceptualization of stress has evolved through both psychological and physiological perspectives. One of the most influential frameworks is the Transactional Model of Stress and Coping, introduced by Lazarus and Folkman [10], which emphasizes that stress arises when perceived demands exceed available coping resources. This model highlights the role of cognitive appraisal, where individuals evaluate a situation as either a challenge or a threat, ultimately influencing their emotional and behavioral responses.

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

On the physiological side, Selye's General Adaptation Syndrome (GAS) describes stress as a biological process comprising three stages: the alarm phase, where the body recognizes stressors; the resistance phase, where adaptive mechanisms are activated; and the exhaustion phase, where prolonged exposure depletes resources and may lead to health complications [11]. Together, these models provide a foundation for understanding how stress manifests in academic environments, where cognitive overload, peer pressure, and high-stakes assessments act as significant triggers.

B. Stress Biomarkers

Biomarkers are measurable indicators that reflect an individual's stress state. These can be broadly categorized as physiological, behavioral, and multimodal. Physiological biomarkers include electroencephalography (EEG), electrocardiography (ECG), heart rate variability (HRV), galvanic skin response (GSR), and cortisol levels [12], [13]. EEG is widely used in stress detection as it directly captures variations in brain activity linked to cognitive and emotional processing [14]. ECG and HRV are reliable indicators of autonomic nervous system balance and are frequently employed to monitor stress-induced cardiac variations [15]. Similarly, GSR reflects sympathetic nervous system activation and is particularly useful for detecting short-term stress fluctuations [16]. On the behavioral side, micro-expressions, speech intonation, eye-gaze dynamics, and body posture provide rich insights into stress responses [17]. Recent multimodal approaches combine these biomarkers, such as fusing EEG with facial expressions or ECG with voice analysis, to achieve higher robustness and accuracy. In classroom environments, where stress is influenced by both internal (cognitive load) and external (teacher–student interaction) factors, multimodal biomarkers offer an effective solution for real-time monitoring.

C. Traditional Stress Detection Methods

Conventional methods of stress detection have primarily relied on self-reported psychological scales and observational approaches. Tools such as the Perceived Stress Scale (PSS), the State– Trait Anxiety Inventory (STAI), and the Depression Anxiety Stress Scales (DASS) remain popular due to their simplicity and ease of administration [18]. However, they are inherently subjective and unable to provide continuous stress monitoring, which limits their applicability in dynamic settings such as classrooms. Teacher-led behavioral observations and manual performance assessments offer additional insights but are prone to bias, inconsistency, and observer fatigue [19]. While sensor-based techniques, including EEG headbands, ECG chest straps, and wristworn GSR devices, provide more objective insights, their intrusive nature often reduces student comfort and natural classroom behavior. These limitations underscore the importance of developing non-intrusive, automated, and scalable methods for stress detection. In this regard, deep learning–based approaches show significant promise, as they are capable of handling multimodal data streams, reducing reliance on subjective inputs, and enabling real-time analysis for educational applications.

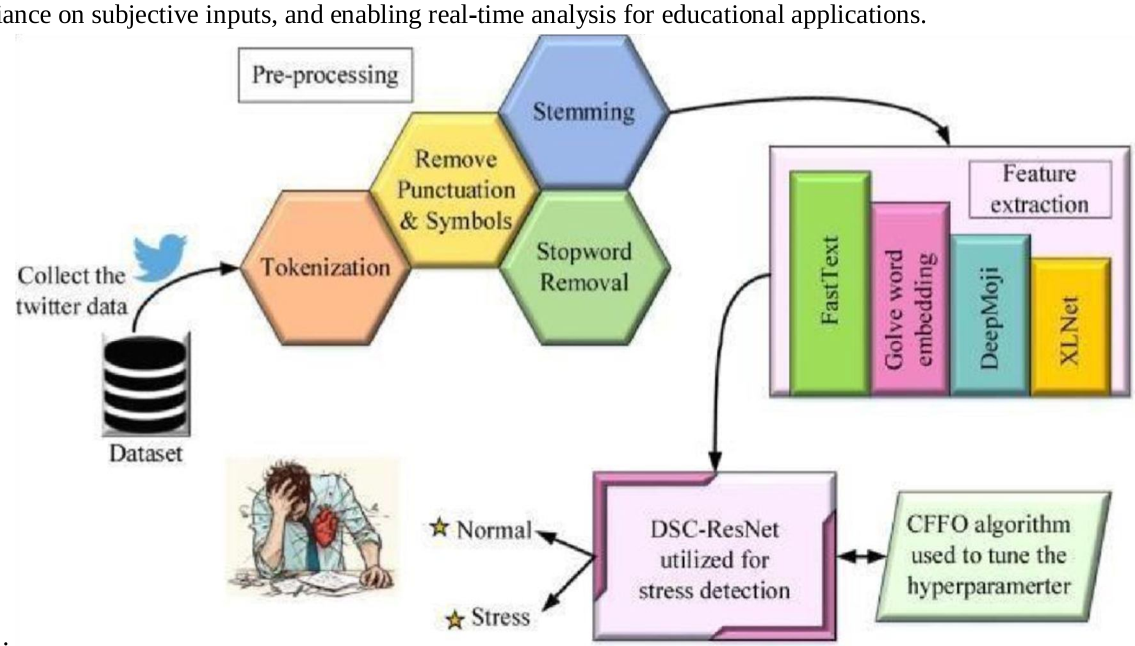


Fig 1 Personalized Stress Detection Technique Using A Deep Learning Model

III. DATA SOURCES AND MODALITIES

The reliability of deep learning–based stress detection in classroom environments is significantly determined by the type and quality of data used for model training and validation. Stress manifests in multiple domains, including physiological, behavioral, and multimodal signals. Each modality captures different aspects of the stress response—physiological measures directly reflect the body's autonomic activity, while behavioral cues provide observable indicators of emotional state. Multimodal approaches aim to combine these perspectives to achieve more robust detection. In this section, we provide a detailed discussion of these data sources, highlighting their characteristics, challenges, and suitability for classroom-based research.

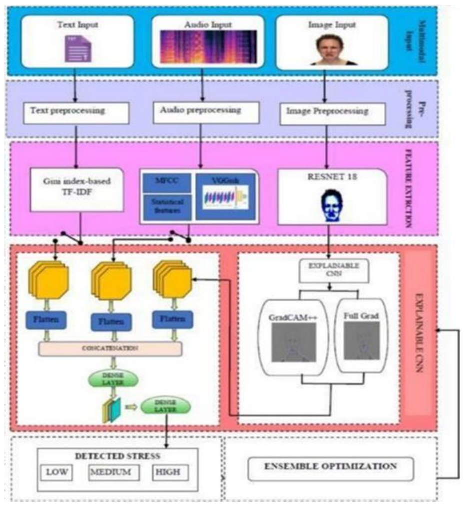


Fig 2 Data Sources And Modalities In Stress Detection For Class Room Environment

Multiple data sources and modalities are used to detect stress in a classroom environment, as a single source is often insufficient for accurate results. Data can be categorized as physiological, behavioral, and environmental. The most effective systems use multimodal fusion, combining data from several sources to achieve higher accuracy.

A. Physiological Signals

Physiological signals remain among the most widely studied indicators of stress due to their close connection to underlying neural and autonomic processes. These signals are typically acquired through wearable or non-invasive biosensors, which can capture real-time responses such as EEG (electroencephalography), ECG (electrocardiography), GSR (galvanic skin response), HRV (heart rate variability), and respiration rate. EEG signals, in particular, provide insights into stress-induced alterations in brain activity, with changes in alpha, beta, and theta band power being strongly correlated with cognitive overload. Similarly, ECG and HRV provide sensitive measures of sympathetic and parasympathetic nervous system balance, which are disrupted under stress conditions [20].
Electrodermal activity (EDA), measured through GSR, reflects sweat gland activity controlled by the sympathetic nervous system and has been widely validated as a marker of stress-related arousal [21]. Respiration patterns, including irregular breathing and reduced variability, further contribute to the assessment of psychological tension [22]. While these modalities provide objective and quantifiable measures, their integration into classroom environments poses challenges such as the need for wearable devices, potential discomfort for students, and issues of signal artifacts caused by movement. However, recent advances in lightweight wearable sensors and unobtrusive biosignal monitoring devices are improving feasibility for classroom applications.

B. Behavioral Indicators

Behavioral signals provide an accessible and non-intrusive means to detect stress, making them particularly attractive in educational contexts. These include facial expressions, vocal characteristics, eye gaze, gestures, and body posture. Stress often induces micro-expressions, reduced smiling, and tense facial movements that can be effectively captured using computer vision algorithms. Deep learning–based facial analysis systems, especially CNNs, have demonstrated the ability to recognize subtle stress-related expressions in real time [23]. In addition to facial cues, vocal features such as pitch, speech rate, and spectral properties provide valuable information about stress levels. Studies have shown that stressed individuals often exhibit higher pitch variability, reduced speech fluency, and uneven energy distribution across frequency bands [24].

These features can be extracted through automated speech recognition pipelines and combined with acoustic emotion recognition techniques for classroom stress analysis.

Body language and posture also reflect stress, where behaviors such as slumping shoulders, restlessness, or repetitive movements signal cognitive and emotional strain [25]. Eye-tracking technologies provide further insight into stress-related attentional changes, with prolonged fixation or avoidance behavior often linked to anxiety and cognitive overload [26]. Compared to physiological signals, behavioral indicators are less intrusive, as they can be collected using cameras and microphones already present in classrooms. However, environmental variability (e.g., lighting, background noise) and cultural differences in expression remain important challenges in ensuring accuracy and generalizability.

C. Multimodal Fusion

Single-modality approaches often fail to capture the full complexity of stress, as each signal type is influenced by different sources of variability and noise. To overcome these limitations, multimodal fusion strategies combine physiological and behavioral data streams to produce more accurate and reliable predictions. For example, fusing EEG data with facial expression analysis has been shown to significantly improve classification performance compared to using either modality alone [27].

Multimodal fusion can occur at different levels: early fusion, where features from multiple modalities are concatenated before being fed into the model; late fusion, where independent classifiers for each modality are combined at the decision stage; and hybrid fusion, which integrates both strategies to balance feature richness and interpretability [28]. Recent advances in deep learning, particularly CNN–LSTM hybrids and transformer-based architectures, have enabled effective modeling of temporal dependencies across multimodal data streams.

In classroom environments, multimodal frameworks are particularly advantageous, as they can leverage a combination of non-invasive sensors (e.g., cameras, microphones) with optional wearable devices (e.g., smartwatches measuring HRV). Such integration provides a holistic understanding of both internal physiological states and external behavioral cues. Although multimodal systems are computationally demanding and raise concerns about privacy and data management, they represent the most promising direction for developing robust stress detection frameworks in real-world educational settings.

Table I: Comparison of Physiological, Behavioral, and Multimodal Data for Classroom Stress Detection

DATA TYPE	EXAMPLES	ADVANTAGES	LIMITATIONS	SUITABILITY IN CLASSROOM
Physiological	EEG, ECG, HRV, GSR, PPG	Objective and quantifiable; High accuracy in measuring stress response; Less bias than selfreports	Requires wearable devices; May cause discomfort; Possible intrusiveness in natural classroom settings	Useful for research-based monitoring; suitable in controlled classroom studies with consent
Behavioral	Facial expressions, body posture, eye gaze, speech tone	Non-intrusive; Can be captured using cameras/microphones; Reflects real-time affective state	Sensitive to lighting, occlusions, background noise; Privacy concerns; Interpretation may vary	Highly suitable for classroom monitoring; enables continuous, passive stress detection
Multimodal	Combination of physiological + behavioral signals	Provides richer context and higher accuracy; Robust against singlesensor limitations; Captures complex stress dynamics	Computationally intensive; Requires synchronization of modalities; More complex data collection setup	Highly suitable for adaptive learning systems; balances accuracy and contextual understanding

Table I provides a comparative summary of physiological, behavioral, and multimodal data sources, highlighting their respective advantages, limitations, and classroom applicability."

IV. DATA SOURCES AND MODALITIES

Stress detection in classroom environments relies on diverse data sources that capture both internal physiological responses and external behavioral cues. The choice of modality significantly affects the accuracy, feasibility, and ethical considerations of stress monitoring in real-world academic settings. This section outlines the key modalities, including physiological signals, facial and micro-expressions, speech and voice analysis, and multimodal datasets that integrate multiple streams of information for robust stress recognition.

A. *Physiological Signals (EEG, ECG, EDA, HRV)*

Physiological signals provide direct, objective measurements of the body's response to stress. Electroencephalography (EEG) measures electrical brain activity, offering insights into cognitive load and emotional regulation. Electrocardiography (ECG) captures cardiac signals, with heart rate variability (HRV) serving as a well-established biomarker of stress and autonomic nervous system activity. Electrodermal activity (EDA), often measured through galvanic skin response, reflects sympathetic nervous system arousal and has been widely applied in emotion recognition studies. These modalities are reliable for stress detection, but their application in classrooms requires careful consideration of intrusiveness, student comfort, and data privacy.

B. *Facial Expressions and Micro-Expressions*

Facial expressions are strong behavioral indicators of emotional states and can be non-invasively captured through computer vision techniques. Macro-expressions, such as frowning or smiling, provide clear markers of affect, while micro-expressions—brief, involuntary facial movements—offer subtle cues of hidden or suppressed stress. With advances in deep learning and convolutional neural networks (CNNs), facial expression recognition has achieved significant accuracy in classroom-based studies. However, factors such as lighting conditions, occlusion, and cultural variations in expressivity may influence detection reliability. Despite these challenges, facial analysis remains one of the most practical and scalable approaches for monitoring stress in real-time learning environments.

C. *Speech and Voice Analysis*

Speech and vocal characteristics, including tone, pitch, rhythm, and speech rate, carry critical emotional information that reflects stress levels. Under stress, individuals may exhibit changes in vocal pitch, reduced fluency, or increased speech disfluencies. Machine learning techniques such as recurrent neural networks (RNNs) and long short-term memory (LSTM) models have been employed to capture temporal variations in speech features, enabling robust stress classification. In classroom settings, microphones embedded in learning environments can be used to analyze students' speech during interactions, presentations, or discussions. The main challenges include background noise, overlapping voices, and privacy concerns; nevertheless, voice analysis provides a valuable, non-invasive window into students' emotional states.

D. *Multimodal Datasets in Classroom Environments*

While single-modality data sources provide valuable insights, multimodal datasets offer a more comprehensive view by combining physiological and behavioral signals. For example, EEG and ECG data can be integrated with facial expression analysis to enhance stress classification accuracy. Similarly, speech features combined with eye-gaze tracking or posture monitoring provide a richer context of cognitive and emotional states. Several benchmark multimodal datasets, such as DEAP and DREAMER, exist for affective computing; however, dedicated classroomoriented datasets are still limited. The development of large-scale, ethically collected multimodal classroom datasets is crucial to designing adaptive learning systems that can personalize teaching strategies based on students' stress levels.

Table II: Summary of Stress Detection Modalities in Classroom Environments

MODALITY	KEY FEATURES	ADVANTAGES	CHALLENGES	REPRESENTATIVE SOURCES
Physiological Signals	EEG, ECG/ HRV, EDA (GSR), respiration	Objective	Requires wearables;	EEG [1], HRV [20], EDA [14]

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

Facial Expressions & Micro-Expressions	Macro- and micro-level facial cues, action units (FACS/AUs)	Non-intrusive;	Sensitive to lighting, occlusion, cultural expression differences	Compound Action Units [9]
Speech & Voice Analysis	Vocal pitch, mel-spectrogram features, fluency, energy distribution	Non-invasive; Easily recordable via classroom audio	Background noise; Multiple speakers; Privacy and consent issues	Speech OpenSMILE features [21]
Multimodal Fusion Approaches	Integration of physiological + behavioral data using early/late fusion or hybrid deep models	Enhanced accuracy; Robust via complementary signals	High computational complexity; Synchronization issues; Privacy considerations	EmpathicSchool dataset [13], Real-time deep fusion [4]

,

Table II provides a concise comparison of the main data modalities used in classroom stress detection, emphasizing their key features, strengths, challenges, and relevant literature."

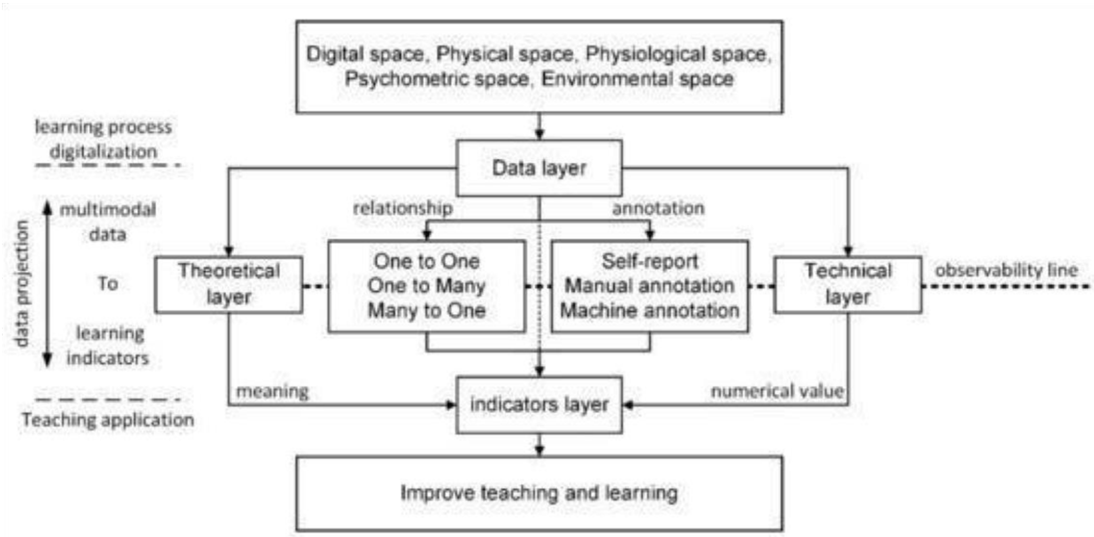


Fig 3. A Conceptual Model Of Multimodal Data Analysis.

The diversity of physiological, behavioral, and multimodal data sources highlights the complexity of stress detection in classroom environments. While physiological signals provide quantifiable measures, facial and vocal cues capture naturalistic behaviors that are essential for real-time monitoring. Multimodal datasets, in particular, enable richer contextual understanding by fusing complementary data streams. However, the heterogeneous nature of these data types requires advanced analytical approaches capable of handling high-dimensional, noisy, and multimodal inputs. Traditional machine learning methods often struggle with such complexity, making deep learning a natural candidate for effective feature extraction, fusion, and classification. Consequently, the following section explores state-of-the-art deep learning models for stress detection, emphasizing architectures specifically designed to leverage the unique characteristics of multimodal classroom data.

V. PREPROCESSING AND FEATURE ENGINEERING

Preprocessing and feature engineering play a crucial role in developing effective deep learning models for stress detection in classroom environments. The raw physiological and behavioral data collected from students, including signals such as electrodermal activity (EDA), electroencephalography (EEG), heart rate (HR), and skin temperature, often contain noise, artifacts, and variability that must be addressed before model training. Signal processing techniques, such as noise filtering using Butterworth or Chebyshev filters, segmentation of continuous signals into epochs, and normalization to standard scales, are commonly employed to ensure that the data is clean, consistent, and suitable for learning algorithms. These preprocessing steps enable the extraction of meaningful information that can accurately represent the underlying stress patterns in students [29].

Data augmentation methods are frequently utilized to address the challenges of limited datasets, which are common in classroom-based stress detection studies. By applying techniques such as time-domain augmentation, including jittering and sliding window methods, or frequency-domain adjustments to modify spectral components, the diversity of the dataset can be increased without additional data collection. Moreover, advanced generative approaches, such as Generative Adversarial Networks (GANs), allow for the synthesis of realistic physiological data that closely resembles actual stress responses, thus improving model generalization and reducing overfitting [30].

Feature extraction is a critical step that transforms preprocessed signals into a set of informative attributes suitable for classification by deep learning models. Time-domain features, including mean, standard deviation, skewness, and kurtosis, provide insights into the central tendency and variability of signals. Frequency-domain features, derived through techniques such as Fast Fourier Transform (FFT), capture the energy distribution across different frequency bands, which is particularly important for analyzing EEG and heart rate variability. In addition, timefrequency domain methods, such as wavelet transforms, allow for the simultaneous analysis of temporal and spectral characteristics, enabling the detection of transient stress patterns that may be overlooked by conventional methods. For enhanced accuracy, multimodal feature fusion is often applied, combining features from multiple physiological and behavioral signals to create a comprehensive representation of a student’s stress state. This careful selection and engineering of features significantly improve the performance and reliability of deep learning models for real-time classroom stress detection [31].

VI. DEEP LEARNING MODELS FOR STRESS DETECTION

The application of deep learning models in stress detection has gained significant attention due to their ability to analyze complex data from various modalities, including visual, auditory, and physiological signals. These models leverage advanced neural network architectures to process and interpret data, providing valuable insights into an individual's stress levels.

A. *Vision-Based Models*

Vision-based stress detection models primarily focus on analyzing facial expressions, body posture, and eye movement to infer emotional states. Convolutional Neural Networks (CNNs) are commonly employed to extract spatial features from images or video frames, capturing subtle facial cues indicative of stress. For instance, models that utilize facial landmarks and regions of interest (ROIs) have demonstrated efficacy in recognizing stress-induced facial expressions. Additionally, the integration of temporal information through Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks allows for the analysis of dynamic facial movements over time, enhancing the model's ability to detect stress-related changes in facial expressions. These vision-based approaches offer a non-intrusive means of monitoring stress, making them suitable for real-time applications in various settings.

B. *Speech-Based Models*

Speech-based stress detection models analyze vocal features such as pitch, tone, speech rate, and energy levels to assess emotional states. Deep learning architectures, including CNNs and Bidirectional LSTM (BiLSTM) networks, are utilized to capture both local and sequential patterns in speech signals.

These models can effectively differentiate between stressed and nonstressed speech by learning complex representations of vocal attributes. The use of pre-trained models like Wav2Vec 2.0 for feature extraction, combined with fine-tuning on domain-specific datasets, has further improved the accuracy of speech-based stress detection systems. Such models are particularly beneficial in scenarios where visual data is unavailable or impractical to collect, providing an alternative means of assessing stress through auditory cues.

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

C. Biosignal-Based Models
Biosignal-based stress detection models focus on physiological data, including Electrocardiography (ECG), Electrodermal Activity (EDA), Heart Rate Variability (HRV), and skin temperature. These signals are processed using deep learning models such as CNNs, RNNs, and Transformer networks to identify patterns associated with stress responses. For example, variations in HRV and EDA have been linked to stress-induced autonomic nervous system activity, and deep learning models can learn to classify these patterns effectively. The advantage of biosignal-based models lies in their ability to provide objective, quantifiable measures of physiological stress, which can be particularly useful in clinical or high-stakes environments where accurate stress assessment is critical.

D. Multimodal Fusion Approaches
Multimodal fusion approaches integrate data from multiple modalities—such as visual, auditory, and biosignal inputs—to create a comprehensive model of stress detection. By combining the strengths of each modality, these models can achieve higher accuracy and robustness compared to unimodal systems. Techniques like early fusion, where features from different modalities are combined at the input level, and late fusion, where individual modality models are trained separately and their outputs are merged, are commonly employed. Intermediate fusion approaches, which combine features at intermediate layers of the network, have also been explored to balance the benefits of early and late fusion. The use of attention mechanisms within these fusion models allows for dynamic weighting of modality contributions, enabling the model to focus on the most informative features at any given time. Multimodal fusion approaches are particularly effective in complex real-world scenarios where stress manifestations are multifaceted and may not be adequately captured by a single modality.

VII. EVALUATION METRICS AND BENCHMARKING

A. Accuracy, Precision, Recall, and F1-Score
In classification tasks, metrics such as accuracy, precision, recall, and F1-score are commonly used to evaluate model performance. Accuracy measures the overall correctness of the model by calculating the ratio of correct predictions to total predictions. However, accuracy can be misleading in imbalanced datasets, where the majority class dominates. Precision and recall offer more nuanced insights: precision indicates the proportion of true positive predictions among all positive predictions, while recall measures the proportion of true positive predictions among all actual positives. The F1-score, defined as the harmonic mean of precision and recall, provides a single metric that balances the trade-off between precision and recall. These metrics are particularly useful in stress detection scenarios, where false positives and false negatives can have significant implications for user well-being and model trustworthiness.

1) Accuracy (Acc),

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

where:

- *TPTPTP = True Positives*
- *TNTNTN = True Negatives* *FPFPFP = False Positives*
- *FNFNFN = False Negatives*

2) Precision quantifies the proportion of correctly predicted positive instances among all predicted positives:

$$\textbf{Prec} = \frac{TP}{TP+FP}$$

3) Recall (or Sensitivity) measures the proportion of actual positives correctly identified

$$Rec = \frac{TP}{TP+FN}$$

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

4) *The F1-score is the harmonic mean of precision and recall:*

$$F1 = 2.\frac{Prec.Rec}{Prec + Rec}$$

These metrics are crucial in classroom stress detection because both false positives and false negatives can significantly impact interventions and student support.

B. Stress-Level Classification vs. Regression

Stress detection models can be broadly categorized into classification and regression approaches. Classification models predict discrete stress levels, such as "low," "medium," or "high," based on input features. These models are evaluated using metrics like accuracy, precision, recall, and F1-score. In contrast, regression models predict continuous stress levels, providing a more granular assessment of stress intensity. Evaluation metrics for regression models include Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), which quantify the difference between predicted and actual stress levels. The choice between classification and regression approaches depends on the specific requirements of the application and the nature of the available data.

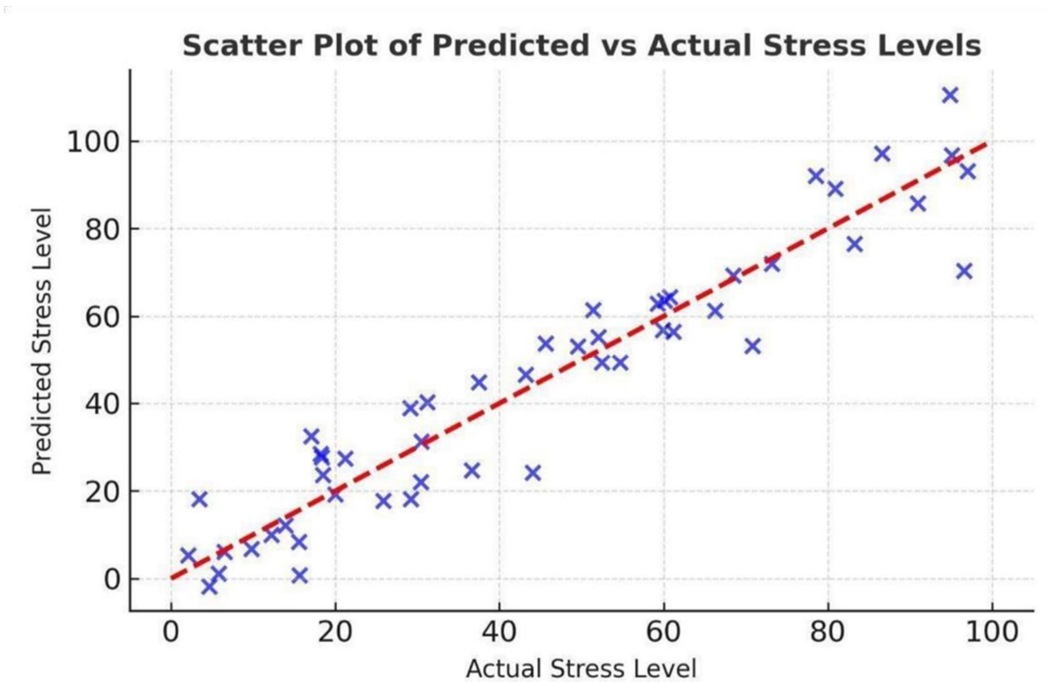


Fig 5 PREDICTED VS. ACTUAL STRESS LEVELS

C. Benchmark Datasets Used in Classroom Studies

Benchmark datasets serve as the foundation for designing, training, and validating deep learning models in stress detection research. Their importance lies not only in providing a structured basis for experimentation but also in enabling reproducibility, fair performance comparison, and the establishment of standardized evaluation practices. In the context of classroom environments, datasets that capture physiological, behavioral, and environmental modalities are especially relevant, as they reflect the multifaceted nature of stress among students. One widely used dataset is the Affective ROAD dataset**,** which provides continuous stress-related recordings collected via wearable sensors such as electrocardiogram (ECG) and galvanic skin response (GSR). Its temporal richness makes it suitable for both classification and regression-based tasks, allowing models to predict stress levels dynamically rather than only in discrete categories.

Another notable resource is the Stress ID dataset, which integrates multimodal data, including facial expressions, audio recordings, and physiological signals. This multimodality supports the development of hybrid deep learning models capable of capturing both observable behavioral cues **and** hidden physiological responses, thereby improving detection accuracy in real-world classroom conditions.

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

Additionally, the Student Stress Monitoring Datasets available on Kaggle provide comprehensive data that extend beyond physiological measures. These datasets encompass psychological, academic, social, and environmental stress factors**,** reflecting the diverse challenges students face in educational environments. Such richness enables the development of holistic frameworks that not only detect stress but also model its contextual triggers.

The use of benchmark datasets in classroom-focused studies offers several advantages.

- Compare different deep learning architectures under controlled and standardized conditions.
- Identify critical features (e.g., heart rate variability, speech pitch, micro-expressions) that strongly correlate with stress responses.
- Develop generalizable models that can be validated across institutions and learning environments.
- Address domain-specific challenges such as data imbalance, multimodal fusion, and real-time deployment feasibility.

By leveraging these datasets, the research community can accelerate progress toward creating robust, scalable, and ethically sound stress detection systems for classroom applications. However, ongoing efforts are needed to expand dataset diversity, incorporate longitudinal recordings, and address privacy issues, ensuring that future benchmarks reflect both the complexity of stress and the practical requirements of smart educational environments.

Table III. Benchmark Datasets Commonly Used For Stress Detection In Classroom Studies

DATASET	MODALITIES CAPTURED	KEY FEATURES	APPLICATIONS IN
Affective ROAD	Physiological signals (ECG, GSR, respiration, skin temperature)	Continuous stress recordings using wearable sensors; supports time-series analysis	Suitable for classification and regression tasks; real-time stress monitoring
Stress ID	Multimodal data (facial expressions, speech/audio, ECG, GSR)	Combines behavioral and physiological responses; multimodal fusion possible	Development of hybrid models for accurate detection in real-world conditions
Student Stress Monitoring (Kaggle)	Psychological, physiological, academic, social, and environmental factors	Comprehensive dataset covering multiple stress dimensions	Holistic stress modeling; analysis of contextual triggers in classroom environments
WESAD (Wearable Stress and Affect Detection)	Physiological signals (ECG, EDA, EMG, respiration, body temperature, accelerometer)	High-quality multimodal dataset collected in lab and semi-realistic settings	Widely used benchmark for stress classification and model comparison
DEAP (Database Analysis using Physiological Signals)	EEG, peripheral physiological	Includes emotional states with annotated labels	Useful for emotion recognition tasks, often extended to stress analysis in learning scenarios

VIII. APPLICATIONS IN CLASSROOM ENVIRONMENTS

The deployment of deep learning–based stress detection frameworks in classroom environments is emerging as a transformative approach to modern pedagogy. Unlike traditional teaching methods, which rely largely on uniform instructional delivery, these intelligent systems enable the development of personalized, adaptive, and data-driven learning ecosystems. By continuously monitoring physiological, behavioral, and multimodal signals, stress detection applications provide actionable insights into student well-being and learning capacity. The resulting benefits include real-time stress assessment, adaptive instructional mechanisms, and early intervention strategies, all of which contribute to improved academic outcomes and emotional support within educational settings.

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

A. Real-Time Stress Monitoring

Real-time stress monitoring focuses on the continuous collection and interpretation of student data to identify variations in stress levels as they occur. Physiological signals such as electro dermal activity, heart rate variability, and respiration patterns are commonly used, in combination with behavioral cues such as facial expressions, vocal tone, and body posture. The integration of these multimodal indicators provides a comprehensive representation of student stress responses.

Deep learning models play a central role in enabling accurate and timely detection. Convolutional Neural Networks are particularly effective in extracting spatial features such as microexpressions and subtle visual cues from facial images, while Long Short-Term Memory networks excel in capturing the temporal dependencies present in sequential data like heart rate fluctuations and speech variations. By combining these models, stress detection frameworks are capable of producing robust classifications across different stress levels.

The general process involves several stages: first, data acquisition is achieved through wearable sensors, video cameras, and microphones embedded in classroom environments. Next, preprocessing ensures signal quality by removing artefacts and normalizing data across modalities. This is followed by feature learning, where deep neural networks automatically extract relevant characteristics from raw input streams. In the classification stage, stress levels are categorized into low, moderate, or high using machine learning classifiers or ensemble techniques.

Finally, the system provides real-time feedback, allowing teachers and administrators to intervene through strategies such as pacing adjustments, relaxation breaks, or personalized instruction. A typical system pipeline can be expressed as:

This approach allows teachers to detect elevated stress states during lectures or assessments and intervene before performance

$$S_t = f_\theta(X_t^{phys}, X_t^{behav})$$

where S_t represents the estimated stress level at time t, X_t^{phys} denotes physiological signals, X_t^{behav} indicates behavioral cues, and f_θ is the deep learning function parameterized by θ.

deteriorates. Studies have shown that wearable-based stress monitoring systems integrated with machine learning can achieve accuracies exceeding 85% in real-time scenarios [32].

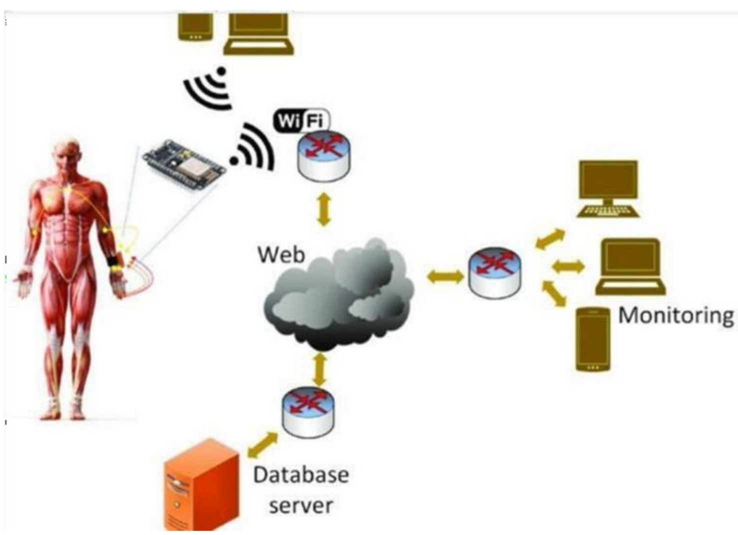


Fig 6. Real-Time Stress Monitoring System Pipeline

B. Adaptive Learning Mechanisms

Adaptive learning mechanisms represent one of the most promising applications of stress detection in classroom environments. These mechanisms aim to tailor the pace, style, and delivery of instruction based on the stress levels and emotional states of students. By integrating stress detection systems into smart classrooms, teaching strategies can shift from a uniform approach to a more dynamic and personalized mode of instruction. This transformation allows educators to respond not only to cognitive needs but also to the psychological readiness of learners.

In practice, adaptive learning begins with the real-time identification of a student's stress state through multimodal inputs such as physiological signals, behavioral patterns, and speech characteristics. When elevated stress levels are detected, the system can automatically adjust instructional parameters, for example by slowing down the pace of lecture delivery, simplifying the complexity of presented material, or incorporating interactive and engaging activities to reduce cognitive load. Conversely, when students exhibit low stress and high engagement, the system can accelerate instruction or introduce advanced material to maintain an optimal learning curve.

Deep learning models, including Convolutional Neural Networks, Long Short-Term Memory networks, and Transformer-based architectures, play a crucial role in enabling adaptive mechanisms. These models capture both short-term fluctuations and long-term trends in stress responses, thereby allowing the system to make accurate adjustments in real time. Over time, the system can also learn from historical data, creating personalized stress-response profiles for individual students and refining interventions to suit their unique learning patterns.

The benefits of adaptive learning mechanisms extend beyond academic performance. They contribute to the creation of inclusive classrooms where students with varying stress tolerances and coping mechanisms receive equal opportunities for success. Moreover, such systems promote resilience by encouraging healthy stress management strategies, ensuring that learning remains both effective and sustainable.

Formally, the adaptive mechanism can be modeled as:

$$A_t \,\square\, f\,(S_t, P_t, B_t, E_t \square)$$

Where: A

- t → Adaptive learning adjustment at time t
- S_t → Stress index derived from physiological and behavioral signals
- P_t → Performance metrics (e.g., task accuracy, engagement score)
- B_t → Behavioral cues (e.g., facial expression, posture, interaction patterns)
- E_t → Environmental/contextual factors (e.g., workload, classroom conditions)
- → Model parameters that are continuously updated

The optimization objective can be written as: $\lim L(A_t, A_t \square) \;\square\square\; R(\square)$

□

$L(\;)$ □ Loss function measuring deviation between predicted adaptive response A_t and desired response A_t $R(\;)$ □

□ Regularization to prevent overfitting and ensure fairness

□□ Weighting factor balancing adaptation accuracy and stability

Such personalized interventions foster a balance between cognitive load and learning efficiency, thereby reducing academic anxiety and improving long-term retention [33].

Table IV: Examples Of Adaptive System Responses Under Varying Stress Levels

STRESS LEVEL	OBSERVED INDICATORS (PHYSIOLOGICAL/BEHAVIORAL)	ADAPTIVE SYSTEM RESPONSE	EXPECTED OUTCOME
Low Stress	Stable heart facial expressions, active participation	Maintain baseline content delivery	Sustained engagement and normal learning pace
Moderate Stress	Slightly elevated heart rate, reduced eye contact, slower task completion	Adjust task difficulty, provide supportive prompts	Reduced anxiety, improved task completion
High Stress	Rapid heart rate, frowning, disengagement, frequent errors	Trigger relaxation strategies (short breaks, mindfulness cues), simplify content	Stress reduction, regained focus, improved performance
Critical Stress	Strong physiological arousal, withdrawal from activity, refusal to participate	Immediate alert to instructor, peer-support or counseling referral	Prevention of intervention for well-being

C. Early Intervention and Support for Students

Early interventions form a critical component of stress detection applications in classroom environments. The primary objective of such interventions is to provide timely support to students who exhibit signs of elevated stress, thereby preventing prolonged cognitive overload, anxiety, or burnout. By identifying stress at an early stage, educators can apply corrective strategies before the condition negatively impacts academic performance or psychological well-being.

The process of early intervention begins with continuous monitoring of stress signals and the subsequent computation of stress levels. When the detected stress index surpasses a predefined threshold, the system generates an alert that can be directed either to the student or to the instructor. For instance, if stress levels are consistently high during a specific learning activity, the system may recommend relaxation techniques, short breaks, or the introduction of collaborative tasks to reduce pressure. Similarly, teachers may receive feedback to adjust instructional complexity or provide additional guidance to struggling students. Stress Detection with Early Intervention:

$$I(S) = \{10 \text{if } SI \geq \theta \text{if } SI < \theta \text{ Where:}$$

- SI = Stress Index (calculated from physiological + behavioral parameters).

□

θ= predefined stress threshold.

- I(S) = Intervention Indicator (1 = intervention triggered, 0 = no intervention).

where HRV represents heart rate variability, EDA denotes electrodermal activity, F_{exp} corresponds to facial expression indicators, and V_{tone} reflects voice modulation features. The coefficients $\alpha, \beta, \gamma, \delta$ are weighting factors assigned based on the relative contribution of each modality in predicting stress.

By applying such formulas in real time, early intervention systems can not only detect heightened stress but also differentiate between temporary fluctuations and persistent stress patterns. This enables targeted support strategies that are both timely and personalized.

Early interventions therefore serve as a preventive mechanism, ensuring that stress does not accumulate to detrimental levels. They foster a supportive learning atmosphere, reduce the likelihood of long-term academic disengagement, and enhance the overall emotional resilience of students.

The integration of deep learning–based stress detection systems into classroom environments represents a paradigm shift in modern pedagogy. Real-time monitoring enables the continuous assessment of students' emotional states, adaptive learning mechanisms personalize the instructional process according to individual stress responses, and early interventions provide timely support to mitigate potential adverse outcomes. Collectively, these applications establish a holistic framework that not only enhances academic engagement but also promotes psychological well-being and resilience among learners. By combining multimodal data analysis with advanced neural architectures, such frameworks can bridge the gap between technological innovation and practical educational outcomes, ultimately contributing to the development of intelligent, responsive, and student-centered classroom ecosystems.

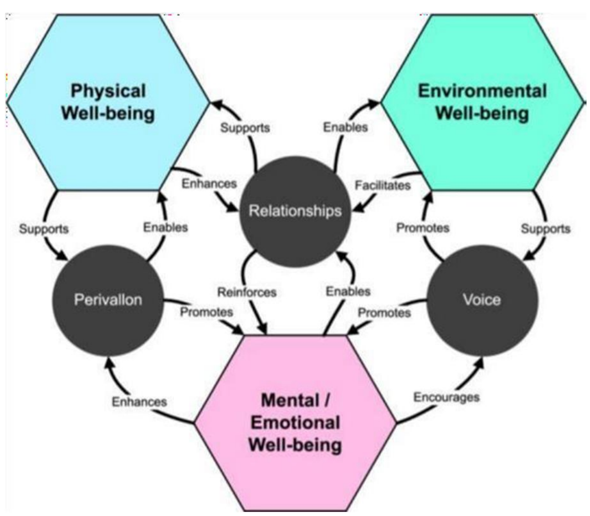


Fig 7: The complexity model of well-being is depicted with the three dimensions of well-being represented as hexagons and influencing factors

These predictions enable schools to provide psychological counseling, peer support, or workload adjustments proactively. Integrating such frameworks into classroom environments has been found to improve not only academic performance but also emotional resilience among students [34].

IX. CHALLENGES AND LIMITATIONS

Although deep learning–based stress detection systems have shown significant promise, their widespread adoption in classroom environments remains constrained by several challenges. These challenges extend beyond technical considerations and include issues related to data acquisition, privacy, ethical acceptability, model generalizability, and hardware deployment. Addressing these concerns is vital to ensure that stress detection frameworks are accurate, equitable, and practical in real-world educational settings.

A. Data Collection in Natural Classroom Settings

The collection of stress-related data in real-world classrooms is inherently more complex than in controlled laboratory environments. Unlike laboratory settings where conditions are standardized, classroom environments are influenced by cultural, social, and situational factors that make stress labeling difficult and often subjective [35]. Physiological signals such as EEG, ECG, and GSR are particularly prone to motion artifacts, device misplacements, and environmental interference when students engage in routine classroom activities. Such inconsistencies reduce the reliability of ground-truth stress labels, thereby limiting the effectiveness of fully supervised learning. Recent research emphasizes semi-supervised learning, domain adaptation, and advanced data augmentation strategies as solutions to counter the scarcity of high-quality annotated classroom datasets [36]. Moreover, multimodal integration—combining physiological, behavioral, and contextual data—has been explored to increase robustness, though at the cost of increased data complexity and resource requirements [37].

B. Privacy and Ethical Concerns

The integration of stress monitoring systems in classrooms introduces significant ethical and privacy challenges. Since these systems often rely on video, audio, or biometric signal capture, the risk of exposing personally identifiable and sensitive data is considerable [38]. In addition, continuous surveillance may trigger discomfort or "surveillance anxiety," paradoxically contributing to elevated stress levels among students. Ethical AI guidelines emphasize the need for explicit consent, data minimization, and strong anonymization before deployment [39]. To address these concerns, privacy-preserving machine learning methods have gained traction. Federated learning allows model training without centralizing raw data, while differential privacy and homomorphic encryption offer further safeguards by mathematically ensuring data confidentiality [40]. Despite these advancements, balancing the need for accurate stress detection with respect for student autonomy and well-being remains an unresolved challenge.

C. Model Generalizability and Bias

Another significant limitation lies in the generalizability of stress detection models across diverse educational environments. Deep learning architectures, such as CNN–LSTM and Transformerbased hybrids, often achieve high performance on specific datasets but suffer when applied to unseen classrooms or culturally diverse populations due to dataset bias [41]. For example, a model trained primarily on a single cultural group may fail to capture stress indicators in students from different ethnic or linguistic backgrounds [42]. This limitation underscores the need for cross-cultural benchmarking, domain adaptation, and transfer learning approaches to improve robustness [43]. Furthermore, bias in training datasets can lead to unfair outcomes, such as disproportionately high misclassification rates for certain student groups. Fairness-aware machine learning frameworks and explainable AI mechanisms are being explored to mitigate these issues and ensure equitable model performance [44].

D. Hardware and Deployment Constraints

Beyond algorithmic limitations, hardware and deployment concerns present substantial barriers to practical classroom implementation. State-of-the-art deep learning models typically require GPUs or cloud-based servers for real-time inference, which are impractical in many educational institutions due to cost, infrastructure, or internet dependency [45]. Wearable sensors, such as EEG headbands and GSR wristbands, though effective for data capture, may cause physical discomfort or distract students, limiting their long-term acceptability in classrooms [46]. To overcome these barriers, research has increasingly shifted toward lightweight deep learning models such as MobileNet, TinyML, and edge AI frameworks that support on-device computation.

These approaches allow stress detection systems to operate with reduced latency and resource consumption, making real-time classroom monitoring more feasible. However, a persistent tradeoff remains between computational efficiency and detection accuracy, which continues to influence model design and deployment strategies [47].

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

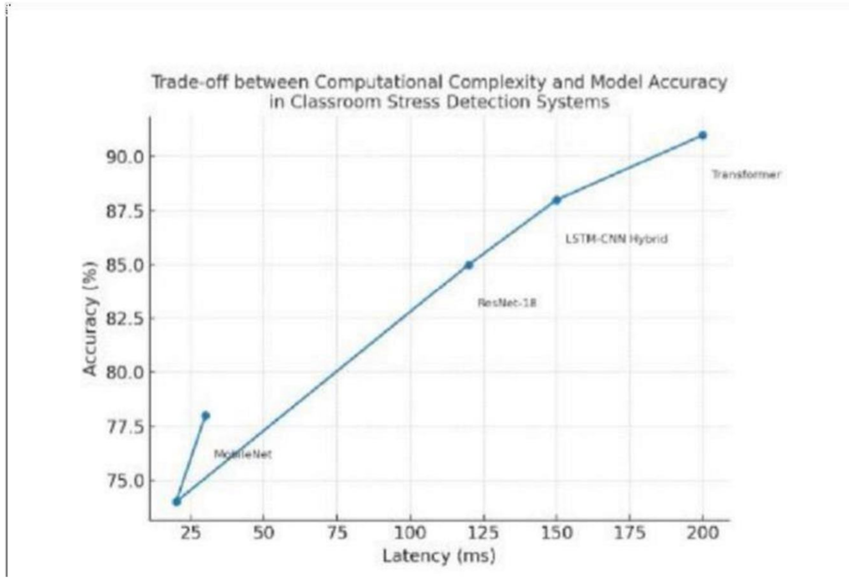


Fig: 8 Complexity and Model Accuracy in class room Stress Detection System

E. Concluding Remarks

In summary, while deep learning–based stress detection systems show strong potential for transforming classroom environments, their practical deployment is hindered by issues of data reliability, privacy and ethics, cross-domain generalizability, and hardware feasibility. Addressing these challenges requires a careful balance between technological innovation and ethical responsibility, ensuring that systems are not only accurate but also transparent, equitable, and sustainable. These limitations also point toward fertile ground for future research, where advancements in explainable AI, privacy-preserving frameworks, and lightweight adaptive models can play a decisive role in overcoming current barriers and enabling large-scale, realworld adoption.

X. FUTURE DIRECTIONS

Despite notable advancements in deep learning–based stress detection, the journey toward real-world classroom adoption remains at an early stage. The classroom setting introduces unique complexities, such as heterogeneous data sources, student privacy concerns, and deployment feasibility, which are not fully addressed by current methods. Moving forward, research must aim to design interpretable, ethical, and resource-efficient systems that support scalable implementation. The subsections below highlight critical directions for future exploration.

A. Explainable Artificial Intelligence (XAI)

A major limitation of deep neural networks in educational stress detection is their “black box” nature, where decision-making processes remain opaque. For teachers, administrators, and policymakers to adopt these systems, models must provide not only accurate predictions but also human-interpretable justifications. Explainable AI (XAI) methods, such as attention-based visualization, saliency maps, and feature attribution techniques, can highlight which physiological signals (e.g., heart rate variability, EEG activity) or behavioral cues (e.g., facial microexpressions, speech prosody) drive model outputs. Beyond trust-building, XAI enables educators to act with greater precision, tailoring interventions based on identified stress triggers. Integrating interpretability modules within CNN–LSTM and Transformer-based frameworks will thus be vital for classroom acceptance and ethical accountability.

For example, attention-based models use a contextual weight distribution:

$$\alpha t = \sum Figure\ 1k = 1Te(ek)exp(et)$$

where αt represents the attention weight at time step , t, and is the learned importance score. These weights highlight which temporal features (e.g., changes in heart rate or facial expression) influence the model’s stress prediction.

B. Wearable and Unobtrusive Technologies

Traditional monitoring tools such as EEG caps or wired sensors, while effective in laboratory contexts, are impractical in dynamic classroom environments. They often disrupt natural learning processes and may cause discomfort for prolonged use. Future research should emphasize unobtrusive technologies, including wearable devices like smartwatches, GSR wristbands, or vision-based monitoring systems capable of capturing subtle behavioral cues without intrusive contact. With advancements in miniaturized sensors and computer vision algorithms, stress indicators can be derived from facial expressions, body posture, or acoustic variations in real time. Such lightweight, student-friendly tools will ensure ecological validity by preserving the natural flow of classroom activities while enabling continuous stress monitoring.

Sfusion = wpSp + wbSb + wcSc where wp, wb, wc are modality-specific weights optimized during training. Such multimodal fusion ensures robust detection across diverse classroom conditions.

C. Federated Learning and Privacy-Preserving Approaches

In educational environments, student privacy is non-negotiable. Conventional centralized machine learning frameworks expose sensitive biometric and psychological data to storage and transmission vulnerabilities. Federated learning offers a paradigm shift by enabling decentralized training across multiple institutions without sharing raw data. When combined with privacypreserving methods such as differential privacy, secure multiparty computation, and homomorphic encryption, it becomes possible to balance data confidentiality with model accuracy. This direction aligns with global ethical AI standards and reduces the risk of surveillance anxiety among students. Implementing federated frameworks will be particularly crucial for multiinstitutional deployments where data governance and compliance are mandatory.

$$minF(w) = i = 1\sum NnniFi(w)$$

where Fi(w)F_i(w)Fi(w) is the local loss for client i, **nin_i***ni*is the dataset size of client$i$, and n is the total number of samples across all clients. This ensures global optimization while preserving local data privacy.

D. Cross-Cultural and Domain-Generalizable Models

Stress manifests differently across cultural, linguistic, and socio-economic contexts, yet many current datasets are narrow in scope, often collected from limited populations. This lack of diversity leads to bias and poor generalizability when applying models to new settings. Future research should prioritize cross-cultural benchmarking datasets and domain adaptation strategies to ensure inclusivity. Approaches such as transfer learning, meta-learning, and adversarial domain adaptation can enable models to adapt across diverse classrooms without significant retraining. Building large-scale, multi-institutional, multicultural corpora will be key to developing stress detection systems that work equitably across varied educational environments.

Bias and limited generalizability remain key barriers. Transfer learning and domain adaptation aim to align feature distributions between source (Ds) and target (Dt) datasets:(Ds) minL(Ds; θ) + λ · MMD(Ds, Dt) where MMD (Maximum Mean Discrepancy) measures the distribution shift, and λ is a trade-off parameter. This improves model adaptability across culturally and contextually diverse classroom environments.

E. Edge AI and Lightweight Architectures

Deploying deep learning models in real-time classroom environments requires balancing computational efficiency with accuracy. High-capacity architectures like Transformers demand significant GPU resources, which are often impractical for school infrastructures. Edge AI solutions, powered by lightweight models such as MobileNet, TinyML, and quantized CNN–LSTM hybrids, offer a feasible alternative. By performing inference directly on local devices, these approaches minimize latency, reduce reliance on external servers, and protect student privacy by limiting data transmission. In the near future, classrooms may adopt hybrid cloud–edge ecosystems, where lightweight stress detection modules operate locally while leveraging cloud resources for periodic retraining and updates. Deploying deep learning models in classrooms requires balancing accuracy and efficiency. Edge AI and lightweight networks reduce computational burden by minimizing parameter size. The optimization goal is:

minL(x; θ) + β · C(θ)

where L(x;θ) is the prediction loss, and C(θ) represents computational cost (e.g., FLOPs, memory). The coefficient β\betaβ regulates the trade-off between accuracy and efficiency.

F. Integration with Adaptive Learning Frameworks

The ultimate value of stress detection lies not in monitoring alone but in fostering personalized learning experiences. Integration with adaptive learning frameworks will allow realtime stress insights to directly inform pedagogical adjustments. For instance, if a system detects heightened stress, teaching platforms could automatically modify lesson pacing, reduce task difficulty, or introduce supportive interventions such as breaks or motivational feedback. Multimodal fusion—integrating physiological, behavioral, and contextual signals—can provide a holistic understanding of student well-being, enhancing the accuracy of adaptive responses. This synergy between stress-aware analytics and adaptive pedagogy has the potential to revolutionize learning by creating supportive, student-centered classrooms. Future systems should connect stress detection outputs with adaptive learning mechanisms that personalize teaching strategies. If y^s\hat{y}_sy^s is the predicted stress probability, the learning difficulty level Ld can be dynamically adjusted as:

$$Ld = Lb - \gamma \cdot y\hat{}s$$

where Lb is the baseline difficulty level, and γ gammaγ is the adaptation rate. This ensures that students experiencing higher stress receive more supportive and less demanding tasks, improving engagement and well-being.

G. Synthesis of Future Directions

In conclusion, the future of deep learning–based stress detection in classrooms depends on developing transparent, unobtrusive, and privacy-aware systems that are robust across diverse populations. Edge AI and federated frameworks will improve real-time feasibility, while integration with adaptive learning systems will ensure that stress monitoring translates into meaningful educational benefits. By aligning technological innovation with ethical and pedagogical goals, future research can transform classrooms into responsive environments that support both academic success and mental well-being.

XI. CONCLUSION

This review has presented a comprehensive analysis of deep learning–based stress detection frameworks in classroom environments, highlighting the interplay between data sources, modeling techniques, and application domains. The discussion underscored the growing importance of multimodal approaches, which integrate physiological, behavioral, and contextual signals to enhance detection accuracy and reliability. Furthermore, recent advances in hybrid architectures—such as CNN–LSTM–Transformer models—demonstrate significant potential for capturing both spatial and temporal dynamics of stress responses in real time.

Applications in classroom settings reveal that stress monitoring can extend beyond passive observation to support real-time interventions, adaptive learning mechanisms, and personalized educational pathways. These systems enable teachers and administrators to better understand student well-being, thereby fostering inclusive, responsive, and emotionally intelligent learning environments. At the same time, the future directions outlined in this review illustrate the need for continued exploration in explainable AI, unobtrusive sensing, privacy-preserving computation, cross-cultural robustness, and lightweight architectures. Addressing these challenges will be critical for scaling deployment and achieving trustworthy integration into educational ecosystems. In conclusion, the synergy of technological innovation and pedagogical sensitivity holds the key to transforming classrooms into adaptive, student-centered ecosystems. By aligning deep learning advances with ethical, cultural, and infrastructural considerations, stress detection research can evolve from experimental studies into real-world solutions. Such progress not onlyenhances academic outcomes but also contributes to the holistic well-being of learners, ensuring that future classrooms are both technologically advanced and emotionally supportive.

REFERENCES

[1] H.-G. Kim, D.-K. Jeong, and J.-Y. Kim, "Emotional stress recognition using electroencephalogram signals with a three-dimensional convolutional gated self-attention deep neural model," Applied Sciences, vol. 12, no. 21, Art. 11162, Nov. 2022, doi: 10.3390/app122111162.

[2] L. Malviya and S. Mal, "A hybrid deep learning framework for EEG-based stress identification," Neural Computing and Applications, vol. 34, no. 22, pp. 17 591–17 604, 2022, doi: 10.1007/s00521-022-07540-7.

[3] P.-F. Zhang, Y. Zhang, L. Gao, and Y. Zhang, "Psychological stress detection from ECG signals using a deep learning attention-based model," Applied Sciences, vol. 11, no. 6, Art. 2848, Mar. 2021, doi: 10.3390/app11062848.

[4] Y. Cho, N. Bianchi-Berthouze, and S. J. Julier, "DeepBreath: Deep learning of breathing patterns for automatic stress analysis using thermal imaging in unconstrained environments," arXiv preprint, arXiv:1708.06026, 2017.

[5] M. Mynoddin, T. Dev, and R. Chakma, "Brain2Vec: CNN–LSTM–attention-based EEG stress detection framework," arXiv preprint, arXiv:2506.11179, 2025.

[6] R. Walambe, P. Nayak, A. Bhardwaj, and K. Kotecha, "Multimodal machine learning for stress detection," arXiv preprint, arXiv:2306.09385, 2023.

[7] J. Agrawal, M. Gupta, and H. Garg, "Early stress detection and analysis using EEG signals within a machine learning framework," IOP Conf. Ser.: Mater. Sci. Eng., vol. 1116, Art. 012134, 2021, doi: 10.1088/1757-899X/1116/1/012134.

[8] H. M. Monisha, R. Sahithi, J. J. Siddarth, and R. A. Kumar, "EEG-based real-time stress detection system using deep learning techniques," Journal of Data Mining and Management, vol. 9, no. 3, pp. 10–21, Oct. 2024.

[9] A. S. Chaudhari and H. Shrivastava, “Survey on EEG-based stress detection using machine learning approaches,” Journal of Electrical Systems, vol. 20, no. 11S, pp. 1–12, 2024. [10] R. S. Lazarus and S. Folkman, Stress, Appraisal, and Coping. New York, NY, USA: Springer, 1984. [11] H. Selye, The Stress of Life. New York, NY, USA: McGraw-Hill, 1956.

[10] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: Analysis of affective physiological state,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175–1191, 2001.

[11] A. Cacioppo, J. Tassinary, and G. Berntson, Handbook of Psychophysiology, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[12] A. Al-Shargie, T. Tang, N. Badruddin, and M. Kiguchi, “Mental stress assessment based on feature level fusion of fNIRS and EEG signals,” IEEE Access, vol. 5, pp. 19889–19896, 2017.

[13] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, “Heart rate variability: A review,” Medical and Biological Engineering and Computing, vol. 44, no. 12, pp. 1031– 1051, 2006.

[14] S. K. Papastefanou, “Galvanic skin response (GSR) in psychology and medicine,” in Proc. 2019 Hellenic Conf. on Artificial Intelligence (SETN), pp. 1–5.

[15] P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.

[16] S. Cohen, T. Kamarck, and R. Mermelstein, “A global measure of perceived stress,” Journal of Health and Social Behavior, vol. 24, no. 4, pp. 385–396, 1983.

[17] C. Spielberger, State–Trait Anxiety Inventory: Bibliography. Palo Alto, CA, USA: Consulting Psychologists Press, 1989.

[18] S. Chikhi et al., “EEG power spectral measures of cognitive workload: A frontal theta index of cognitive load,” Psychophysiology, vol. 59, no. 3, p. e14009, May 2022, doi: 10.1111/psyp.14009.

[19] A. Asif, A. H. Malik, and M. Y. Shaukat, “Human stress classification using EEG signals in response to stressor exposure,” Neurocomputing, vol. 350, pp. 26–41, Feb. 2019.

[20] B. Raufi and L. Longo, “An evaluation of the EEG alpha-to-theta and theta-to-alpha band ratios as indexes of mental workload,” Frontiers in Neuroinformatics, vol. 16, Art. 861967, Jun. 2022, doi: 10.3389/fninf.2022.861967.

[21] T. Y. Wen and S. A. Mohd Aris, “Electroencephalogram (EEG) stress analysis on alpha/beta ratio and theta/beta ratio,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 17, no. 1, pp. 175–182, Jan. 2020, doi: 10.11591/ijeecs.v17.i1.pp175-182.

[22] C.-Y. Chang et al., “Online mental stress detection using frontal-channel EEG recordings in a classroom scenario,” arXiv preprint, arXiv:2405.11394, May 2024.

[23] J. G. Cruz-Garza et al., “EEG-based investigation of the impact of classroom design on cognitive performance of students,” arXiv preprint, arXiv:2102.03629, Feb. 2021.

[24] L. Longo, “Modeling cognitive load as a self-supervised brain rate with electroencephalography and deep learning,” arXiv preprint, arXiv:2209.10992, Sept. 2022.

[25] M. Ahmadi et al., “EEG, pupil dilations, and other physiological measures of cognitive load in a Sternberg memory task,” Multimodal Technologies and Interaction, vol. 8, no. 4, Art. 34, 2024, doi: 10.3390/mti8040034.

[26] B. Raufi and L. Longo, “An evaluation of the EEG alpha-to-theta and theta-to-alpha band ratios as indexes of mental workload,” arXiv preprint, arXiv:2202.12937, Feb. 2022.

[27] A. Ta et al., “Real-Time Stress Monitoring, Detection, and Management in College Students: A Wearable Technology and Machine-Learning Approach,” Sensors, vol. 25, no. 5, p. 2504, 2025.

[28] M. N. A. Khan et al., “Exploring effects of mental stress with data augmentation techniques,” Sensors, vol. 25, no. 2, p. 428, 2025.

[29] X. Ma et al., “The mediating roles of cognitive load and self-efficacy in the relationship between psychological stress and academic performance,” Learning and Individual Differences, vol. 95, p. 102213, 2025.

[30] H. Zhang et al., “Video-based stress detection through deep learning,” Sensors, vol. 20, no. 19, p. 5552, 2020.

[31] X. Ma et al., “The mediating roles of cognitive load and self-efficacy in the relationship between psychological stress and academic performance,” Learning and Individual Differences, vol. 95, p. 102213, 2025.

[32] J. Zhang et al., “Real-time mental stress detection using multimodality: A deep learning approach,” Frontiers in Neuroscience, vol. 16, Art. 947168, 2022.

[33] S. Acharya, “A systematic review of EEG-based automated mental stress detection,” Elsevier, Jan. 2025.

[34] A. Ometov et al., “Stress and emotion open access data: A review,” Int. J. Multimodal Affect. Comput., 2025.

[35] A. Grivet Sébert et al., “Protecting data from all parties: Combining FHE and DP in federated learning,” arXiv preprint, May 2022.

[36] D. Shenoy, R. Bhat, and K. Krishna Prakasha, “Exploring privacy mechanisms and metrics in federated learning,” Artificial Intelligence Review, vol. 58, Art. 223, May 2025.

[37] N. A. Ab. Aziz et al., “Asian affective and emotional state (A2ES) dataset of ECG and PPG for affective computing research,” Algorithms, vol. 16, no. 3, Art. 130, 2023. [40] A. Grivet Sébert, R. Sirdey, O. Stan, and C. Gouy-Pailler, “Protecting data from all parties: Combining FHE and DP in federated learning,” arXiv preprint, May 2022.

[38] S. Bharati, M. R. H. Mondal, P. Podder, and V. B. S. Prasath, “Federated learning: Applications, challenges and future directions,” arXiv preprint, May 2022.

[39] K. Han, Y. Wang, and H. Chen, “Lightweight deep learning models for mobile stress recognition,” IEEE Internet of Things Journal, vol. 9, no. 6, pp. 4201–4214, 2022.

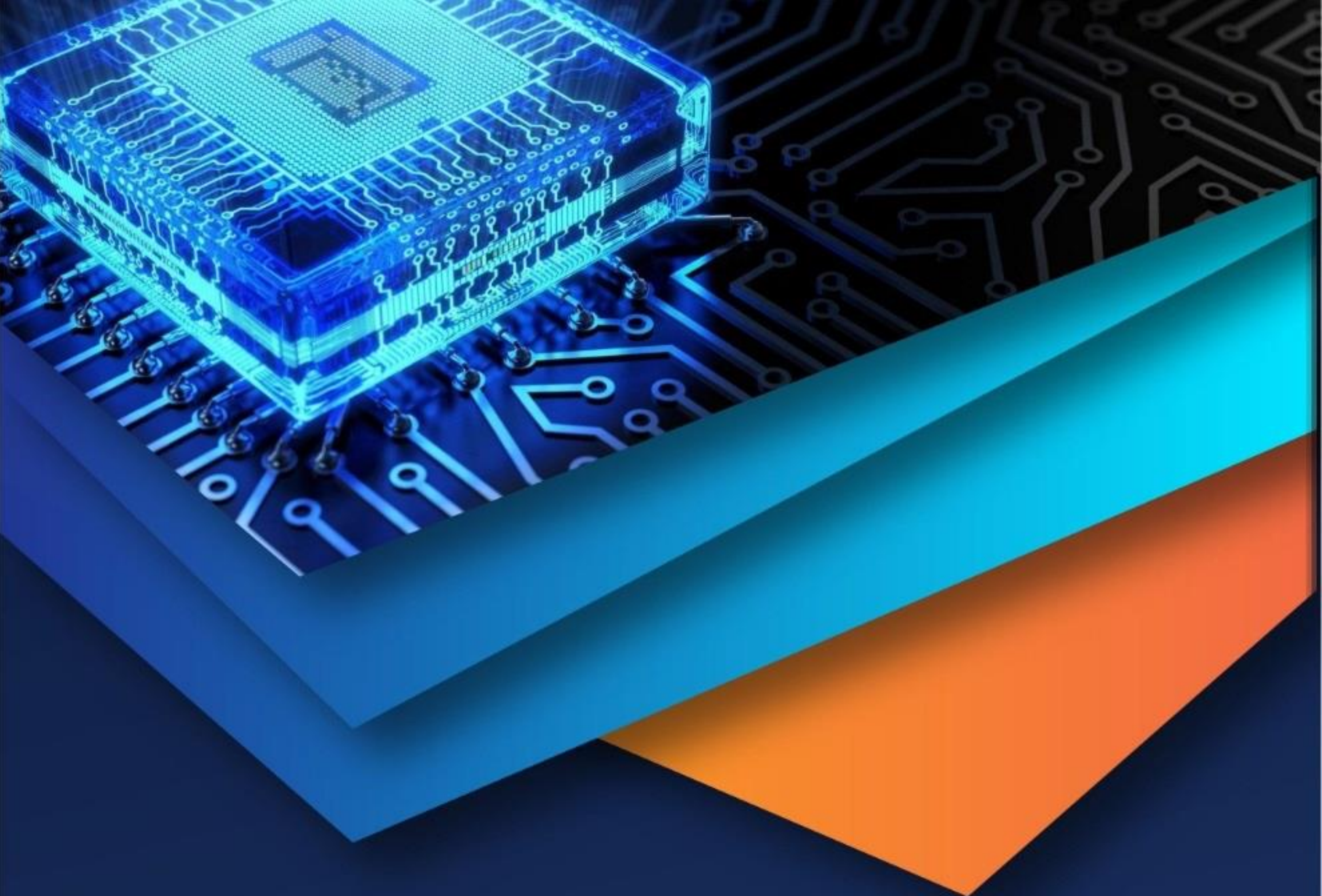
[40] J. Li et al., “Evaluating usability and comfort of wearable stress detection devices in classrooms,” Frontiers in Psychology, vol. 14, Art. 1152345, 2023.

[41] S. Banbury et al., “Benchmarking TinyML systems for edge AI: A case study on stress detection,” in Proc. 2022 Design Automation Conf. (DAC), San Francisco, CA, USA, 2022, pp. 921–928.

[42] S. Kyrou, I. Kompatsiaris, and P. C. Petrantonakis, “Deep learning approaches for stress detection: A survey,” IEEE Transactions on Affective Computing, vol. 16, no. 2, pp. 499– 515, Apr.–Jun. 2025.

[43] V. Hassija, “Interpreting black-box models: A review on explainable AI,” Cognitive Computation, 2024.

[44] H. Liu et al., “Federated learning for privacy-preserving emotion recognition,” IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 9, pp. 4664–4676, 2022.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY