International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

A Review on Stress Detection Using Multimodel for Real-Time Health Monitoring

Ms. Purva Soni
Dept. of Computer Science & Engineering
Prestige Institute of Engineering Management & Research Indore, India

Dr. Dinesh Jain
Dept. pf Computer Science & Engineering
Prestige Institute of Engineering Management & Research Indore, India

***Abstract*—** Stress can be considered one of the most crucial factors that determine the overall physiological and mental wellness of human beings. Serious medical conditions that can result as a consequence of exposure to stress include anxiety disorders, depression, and heart-related ailments. The major drawback of the existing methods to measure stress levels in human beings is that they are largely dependent on self-assessment techniques that are not suitable for continuous tracking. The increased interest in designing a system that can recognize psychological stresses as a consequence of advancements in sensing and machine learning has been accelerated.

This paper offers a thorough look into the latest literature on multimodal stress detection methods that combine the use of physiological sensor information, speech features, and facial image information for real-time health monitoring. The literature is examined for details on the most common datasets, feature extraction algorithms, models of learning, as well as the multimodal fusion methods that are useful in current literature on the topic. Discussion of a comparison between the use of multimodal systems in stress detection and unimodal systems is presented and emphasizes the benefits of multimodal systems over unimodal systems in reliability and generalizability. In particular, potential avenues of research are finally discussed in terms of aiding the development of practical, wearable, and explainable stress detection systems. The purpose of this review is to offer a structured taxonomy and insights towards advancing future works in real-time multimodal stress detection systems.

***Keywords*—** Stress Detection, Multimodal Learning, Physiological Signal, Speech Analysis, Facial Expression Recognition, Health Monitoring.

I. INT DUCTION

Stress is an inextricable part of today's modern life and is a rapidly growing concern for both individual and population health and wellness systems. Today, the rate of urbanization, academic pressure, office stress, and continuous exposure to digital interfaces have caused a marked increase in stress levels in various cross-sections of the population. Chronic and unmanaged stress is known to have a direct correlation to a number of physical and mental ailments such as heart conditions, anxiety, depression, insomnia, and reduced cognitive functions [1], [2]. In view of the serious ramifications, stress identification and continuous surveillance have today assumed importance in prevention and healthcare, mental wellness, and human-computer interaction systems.

Conventional stress evaluation is done by the use of psychology-based measures such as interviews and questionnaires. Though the method is easy and commonly used, the following challenges are involved: the method is prone to a risk of bias based on the previous experiences of the participant; the method is not continuous; and the participant might not be in a position to measure their mental state [3]. Researchers are recently emphasizing the use of objective measures based on data that involves the use of physiological and behavioral parameters.

Objective techniques include stress detection systems based on physiological signals because of their ability to sense internal bodily functions that are directly stressed by stress. Some of the physiological signals that exhibit significant relationships with stress responses include heart rate variability signals, electrodermal activity signals, and respiration signals [4], [5]. Physiological signals are known to be highly prone to artifacts of motion as well as subject variability when considering unconstrained environments.

Another widely considered modality is speech-based stress detection, where stress patterns in speech characteristics like pitch, energy, rate, and spectral properties are identified. Acoustic features, specifically Mel-Frequency Cepstral Coefficients (MFCCs), have been shown to be effective in detecting stress patterns in speech patterns [6], [7]. However, these systems have shown limitations in noisy environments and are affected by language, pronunciation, and emotional ambiguities in speech.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Likewise, facial expression analysis is getting popularity with the development of deep learning methods. Convolutional Neural Networks (CNNs) have made it possible to automatically extract subexpressions about stress, like muscle tension and micro-expressions [8], [9]. However, facial-based stress detection is highly sensitive to lighting conditions, occlusions, head pose variations, and dataset imbalance, which can significantly degrade performance in practical scenarios.

However, the limitations of traditional single-modal stress detection systems have encouraged researchers to work on multimodal frameworks for stress detection. By using the capabilities of more than one modality, such as physiological parameters, speech, and facial expressions, the result becomes more accurate for stress detection [10], [11]. By combining both the physiological parameter and the facial expressions, the missing, noisy, and unreliable patterns in individual modalities can be compensated for in multimodal systems. Recent studies have shown that multimodal systems are better than their single-modal counterparts in terms of accuracy, robustness, and generalizability in the presence of uncertainties in real-world environments [12]-[14].

Apart from the enhanced accuracy, the field of multimodal stress detection has also been aided extensively by machine learning algorithms and, more specifically, deep learning. Methods like ensemble learning, neural networks, or transfer learning have helped significantly in modeling efficiently the intricate, nonlinear relationships existing in the multimodal data patterns [15], [16]. Moreover, fusion methodologies, starting from feature or decision levels or moving towards the hybrid method, are crucial in influencing the system performance or viability [17].

Although considerable advances have been made, a number of issues still linger in the multimodal stress recognition literature. These still-unresolved issues comprise the lack of standardized datasets, privacy issues and ethics, complexity, cross-subject generalization, and issues with real-time implementation, as reported in [18] and [19]. In addition, the existing literature proposes varying conditions concerning modalities, feature extraction, evaluation measures, and experimental conditions, thereby discouraging direct comparison, as pointed out in various research contributions.

This review paper strives to offer a comprehensive and systematic review on multimodal stress detection methods that have been published in existing literature. The current approaches being utilized within existing processes are critically studied based on the types of data being utilized, techniques applied for feature extraction, models followed, and methods employed for fusion. Moreover, an overview on commonly used test databases and evaluation metrics has been presented, followed by a critical review on current limitations. The role presented by this review paper is believed to form a valuable reference for researchers engaged in robust, real-time, and scalable stress detection systems.

II. RELATED WORK

Stress detection studies have developed greatly in the past two decades with the advancements made in sensing, signal processing, and AI. Previous studies can be classified in terms of the modality of data used in stress identification. This section will briefly describe the previous studies on stress detection in terms of both unimodal and multimodal studies.

A. Physiological Signal–Based Stress Detection

Physiological signals have been some of the oldest as well as most popular sources used in objective stress detection, as they directly relate to the autonomic nervous system. Some of the signals commonly used in these studies include heart rate (HR), heart rate variability (HRV), electrodermal activity (EDA), respiration rate, and skin temperature. Such signals relate to the internal body functions that cannot be easily voluntarily controlled, and thus they act as good sources of stress detection.

Healey & Picard proved that HRV-based features could be used effectively to differentiate stress and non-stress conditions in real-world driving environments through machine learning classifiers [8]. Similarly, Rigas et al. used Deep Neural Networks for physiological signals and achieved better results for stress classification compared to existing techniques [26]. Wearable sensor–based stress monitoring systems have also gained popularity due to their portability and continuous monitoring capability [14], [18].

Despite their effectiveness, physiological-based approaches face several challenges. The signal quality of the measures is highly sensitive to motion artifacts, sensor placement, and individual differences in physiological responses [4]. Furthermore, with long-term monitoring come additional concerns with battery life, user comfort, and data privacy. These deficiencies make physiological stress detection systems difficult to scale out under unconstrained environmental conditions.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

B. Speech-Based Stress Detection

Speech analysis has been an emerging mode of detecting stress in a non-invasive and intuitive manner, where variations in vocal characteristics are often typical of the presence of stress. Features to be considered, usually explored in the literature, include pitch, intensity, speaking rate, jitter, shimmer, and spectral coefficients.

Among spectral features, MFCCs are the most widely used as they can capture the stress-induced variations in speech production, according to [10]. Kim and Park utilized MFCC features with neural network classifiers and reported high stress recognition accuracy provided the recording conditions were controlled [5]. Most of the recent works have employed deep learning architectures, such as convolutional and recurrent neural networks, which automatically learn stress-related speech representations [11], [17].

While speech-based detection of stress has excellent performance in controlled laboratory conditions, its robustness decreases under real-world conditions. Environmental noise, linguistic content, accent variations, and emotional overlap can significantly affect classification accuracy [7]. Additionally, speech data may not always be available in continuous monitoring applications, limiting its standalone applicability.

C. Facial Expression–Based Stress Detection

Facial expression analysis relies on the utilization of visual cues to identify stress-related emotional and cognitive states. Consequently, the development of computer vision technology and deep learning has led to the widespread use of Convolutional Neural Networks to identify facial expressions of stress as the state-of-the-art method of facial stress recognition [9], [12].

Huang et al. proved that CNN-based models succeeded in perceiving delicate muscle contraction activities in the human face that relate to stress, and the results were encouraging on benchmark datasets [6]. Transfer learning with VGGNet and ResNet architectures has also been experimented with for better performance on small datasets regarding facial stress recognition [19].

However, facial expression-based stress recognition is confronted with the following challenges: Lighting variations and occlusions (e.g., from glasses or masks) can negatively affect the performance of the system [8], [24]. Further, the facial expression of stress can be subtle and can be confounded with other kinds of emotions.

D. Multimodal Stress Detection Approaches

In an attempt to address limitations in unimodal systems, there has been more attention towards multimodal approaches to stress detection. In multimodal systems, an attempt is made to utilize a variety of sources like speech, facial expressions, and other behavioral characteristics [7], [13].

Poria et al. pointed out that the results of multimodal affective computing always surpass those of single modality systems by exploiting inter-modal correspondences [7]. Chen et al. illustrated the effect of integrating wearable sensor information with behavior cues on improving the accuracy and quality of stress recognition tasks [14]. In multimodal approaches, fusion techniques are vital and depend on three types: feature level fusion, decision level fusion, and hybrid methods [16], [25].

Feature-level fusion involves combining features extracted from various modalities, but this process requires precise synchronization. Another type, decision-level fusion, relates to combining decisions or predictions from separate models designed specifically for individual modalities; this method results in a more robust approach, especially when considering environments where some modalities are absent or unreliable [16].

E. Summary and Research Gaps

Although great progress has been achieved, there are still some existing gaps regarding each mentioned aspect of stress detection studies. Most proposed studies have been carried out using controlled data without tests in real conditions, which surprisingly has little or no comparable use regarding standardized data sets or test procedures among studies [3], [15]. Issues linked to imbalance in data, cross-subject generalizations, the maintenance of privacy, or real-time execution have been open topics [18], [22].

These limitations emphasize the importance of comprehensive reviews that have a structured analysis of existing approaches and give guidance for future studies. This review will address these limitations by structuring the existing work on the basis of modality, learning, and fusion, thus allowing for a comprehensively structured analysis of the current state of art in the field of multimodal stress detection.

III. TAXONOMY OF STRESS DETECTION TECHNIQUES

In order to systematically examine the variety of approaches available for stress detection described in the literature, it is necessary to develop a comprehensive taxonomy of existing solutions. A study based on taxonomy assists in identifying major trends in the literature, and it assists in exploring

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

unresolved issues. A taxonomy-based study is beneficial, unlike comparison studies, in that it facilitates systematic categorization and assists in developing a novel system. For the purpose of this study, a literature review has revealed that stress detection methods can be classified along four primary axes.

A. Classification Based on Data Modality

Also, the most basic aspect of stress detection systems is the type of data modality employed in determining stress levels. On the basis of data modality, the existing systems can be classified into unimodal and multimodal systems.

For unimodal systems, there would be a sole source for data input, including data from human physiology, spoken words, and face expressions. The approach using human physiology would emphasize internal human physiological responses like heart rate variation, electrodermal responses, and respiration patterns [4], [20]. The approach would utilize stress patterns for spoken speech features, while others would emphasize face expressions based on facial muscle tensions [6], [9]. While unimodal systems are easier to develop and distribute, they are handicapped by the weaknesses of the chosen modality.

Multimodal systems incorporate multiple sources of data, each capturing complementary features regarding the stress condition. The combination of physiological signals with speech and facial behavioral cues leads to a more holistic representation of stress and has been shown to perform better in terms of robustness and accuracy than unimodal approaches [7], [13]. Most recent research trends clearly indicate a shift toward multimodal frameworks for stress detection on account of superior performance in practical scenarios.

B. Classification Based on Feature Extraction Techniques

Feature extraction is a very important aspect in stress detection because the quality of features is paramount in model performance. Features in various modalities have vastly different techniques for feature extraction.

For physiological signals, typical features extracted include time domain features like average heart rate and RMSDD, spectral features like LF/HF ratio, and nonlinear features representing autonomic nervous system activity dynamics [5], [18]. For speech-pattern-based stress recognition, common features include acoustic features like MFCCs, pitch, energy, and formants because of their susceptibility to express stress-induced modifications in the speech pattern itself [6], [10]. For facial images, common features include handcrafted features like facial landmarks or features obtained by learning dynamics of convolutional features of the neutral facial expressions automatically by convolutional neural networks [12], [19].

These days, there is an increasing trend in preferring deep feature extraction methods, in which features are learned automatically by neural networks from the original data with minor preprocessing. Though deep features have better accuracy, their requirement of large amounts of data and computational power might restrain their usage in limited settings.

C. Classification Based on Learning Models

Stress detection systems can be found to use a variety of machine learning and deep learning algorithms depending on the modality and application needs. Conventional machine learning algorithms such as Support Vector Machines, Random Forest, k-Nearest Neighbor, and Logistic Regression have been widely used for stress detection in both physiology and speech modalities [8], [9], [21].

Owing to recent advances in deep learning techniques, models built upon neural networks have gained widespread adoption. Convolutional Neural Networks are typically employed for face image analysis tasks, whereas recurrent and feedforward neural networks are used for analyzing time-series characteristics of speech and physiological signals [17], [26]. Various ensemble learning models have also been proposed to combat bias in models and ensure proper generalization.

Although Deep Learning Models tend to perform better compared to traditional methods, the easy understanding and high computational complexity of Deep Learning Models can be pointed out as important challenges.

D. Classification Based on Fusion Strategy

The fusion strategy is a fundamental element of multimodal stress detection systems. The fusion techniques define the manner in which modalities are fused to obtain the result for stress prediction.

Feature-level fusion involves combining features from different modalities into a joint feature vector for the classification process. Though it can harness links among several modalities, it calls for high levels of synchronization and identical features, none of which are readily available [25]. Decision-level fusion makes use of the outputs generated by separate classifiers for each modality via methods like majority voting, weighted voting, and rule-based systems [16].

Hybrid fusion models integrate both feature-level and decision-level fusion. Of all hybrid fusion models, decision-level fusion

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

seems more appropriate for real-world stress recognition, since this method still allows a system to function when there are missing or unreliable modalities.

E. Summary of Taxonomy

In light of the above discussion, techniques for stress detection can be categorized systematically along various dimensions.

Table I. Taxonomy of Stress Detection Techniques Reported in Literature

Study	Modality	Features Used	Learning Model	Fusion Strategy	Key Observation
Healey & Picard [8]	Physiological	HRV features	SVM	Unimodal	Effective in controlled settings
Kim & Park [5]	Speech	MFCC	Neural Network	Unimodal	High accuracy in clean audio
Huang et al. [6]	Facial	CNN features	CNN	Unimodal	Sensitive to lighting
Poria et al. [7]	Multi	Audio + Visual	Deep Learning	Feature-level	Improved robustness
Chen et al. [14]	Multi	Sensor + Behavior	Ensemble	Decision-level	Works with missing data

Table I illustrates a systematic taxonomy of exemplary stress detection methods presented in the literature. It indicates the shift in stress detection systems from unimodal systems to multimodal systems and how decision-level fusion is gaining popularity for real-time implementation because of its flexibility and ability to handle partial modalities effectively.

This taxonomy helps to understand the existing body of research in an organized manner while emphasizing the rising need for multimodal, learning-based, and elastic stress detection approaches.

IV. DATASETS AND EVALUATION METRICS

Dataset and protocol play a vital part in building the effectiveness and validation of the stress detection systems. This is due to the subjective nature of a stress state, which can be benefitted and affected by the data used. Current literature on stress detection systems utilizes a broad variety of datasets, including differences in terms of modality, environment, diversity, and protocols.

A. Physiological Stress Datasets

Datasets related to physiological stress can be acquired through the use of wearable or laboratory-grade sensors, and they can be in the form of heart rate, heart rate variability, electrodermal activity, respiration, and skin temperature signals. Some popular datasets include WESAD, SWELL, and other wearable-related stress detection datasets [14], [18], [21]. These are normally conducted in a controlled manner through stress-related tasks such as public speaking, mental calculations, and time constraints.

Although physiological measures present an objective index of internal stress responses, they can be prone to the problems of subject dependency and a lack of ecological validity [4]. Individual differences in the location of the sensors on the body, level of physical activity, and personal physiological baselines can add noise and impede generalization across subjects [4]. Moreover, many datasets exhibit class imbalance, where non-stress samples significantly outnumber stress samples.

B. Speech-Based Stress Datasets

Datasets of speech samples for stress detection include audio recordings of speech conducted in a neutral state and a state of stress. Speech corpora can be gathered in a lab setting or harvested from act and semiNatural spoken corpora. Speech samples from the corpora are normally annotated according to the speech condition [6], [10].

Speech Corpora contain rich emotional and cognitive information, but they are prone to environmental noise, linguistic information, and audio quality. Also, small size of stress-induced Speech Corpora is common, and this is inadequate for the application of deep learning and might lead to overfitting problems in models [11].

C. Facial Image and Video Datasets

Datasets for facial expressions under stress are image or video files recorded while engaged in stressful tasks. Data is annotated and deduced through experiment setting [9], [12]. Modern sets make use of deep learning-friendly formats that provide both stressful and non-stressing classes.

Despite their utility, facial datasets tend to be imbalanced, less diverse, and vulnerable to environment conditions such as lighting, occlusions, and pose variations. Facial stress detection is one of the most challenging domains within facial detection tasks due to the aforementioned issues, among others, according to reference [19].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

D. Summary of Commonly Used Stress Detection Datasets

To provide a structured overview, Table II summarizes commonly used datasets across different modalities, along with their characteristics.

Table II. Summary of Representative Stress Detection Datasets

Dataset Name	Modality	Data Type	Environment	Key Characteristics	Limitations
WESAD [14]	Physiological	HR, HRV, EDA	Controlled	Multisensor wearable data	Subject-dependent
SWELL [21]	Physiological	HRV, EDA	Office-like	Workload-induced stress	Limited subjects
EMO-DB [10]	Speech	Audio	Acted	High-quality recordings	Acted emotions
SUSAS [11]	Speech	Audio	Simulated stress	Stress-specific speech	Noise sensitivity
FER-based datasets [9]	Facial	Images	Controlled	CNN-ready data	Class imbalance

The comparison of the dataset in Table II draws attention to the point that a majority of MDR studies take place in controlled and semi-controlled environments. Though these datasets help in comparison, they do not encompass the complexity associated with reality-based stress scenarios and thus emphasize the importance of diverse and reality-based environments.

E. Evaluation Metrics for Stress Detection

Evaluation metrics are needed in order to evaluate the efficiency of the stress detection system. Accuracy is the most commonly evaluated metric in the existing systems. But it is not an appropriate metric in cases where there is an imbalance in the class [3]. As a result, other metrics such as precision, recall, and F1 measures are often employed.

Precision is measured by the ratio of correctly predicted stress samples to the total predicted stress samples, and recall is a measure of how well the system is able to identify the true stress samples. The F1 measure offers a harmonic combination of precision and recall, especially in imbalanced datasets [15].

Certain studies have shown confusion matrices for the assessment of classification performance for stress and non-stress classes. In multi-modal systems, the assessment of results gets complicated depending on the modality available, and some studies have followed modality-wise and fusion-level assessment procedures [16].

F. Discussion on Dataset and Evaluation Challenges

Though a lot of work has been conducted, the lack of common datasets and assessment methods is still one of the biggest obstacles in stress detection studies. Various label assignment schemes, participants, sensor set-ups, and experimental conditions introduce complexity in comparing the results of studies directly [18], [22]. Moreover, concerns about privacy and ethics restrict the accessibility of large-scale real-world datasets related to stress.

V. MULTIMODAL FUSION STRATEGIES FOR STRESS DETECTION

Multimodal Stress Detection Systems are dependent upon not only choosing the right data modalities but also upon sophisticated fusion techniques that combine data from different sources. Data fusion is an extremely crucial aspect of judging the level of robustness, scalability, and practicability of a Stress Detection System. Currently, different Stress Detection Systems mostly make use of either Feature-Level Fusion, Decision-Level Fusion, or both.

A. Feature-Level Fusion

Feature-level fusion involves the combination of the features derived from the different modalities into a single joint feature vector prior to the classification process. In the case of stress recognition tasks, this means the combination of the physiological feature vector (e.g. HRV measures) with the acoustic feature vector (from speech) and the visual feature vector (from facial images) [7], [25].

This allows the learning model to directly capture the correlations between the modalities and has proved to be accurate in controlled experiments [13]. However, the requirement for temporal synchronization between the modalities and the need for the features to have the same scaling factors make it unsuitable for stress monitoring in most cases [16]. Moreover, the presence of missing or erroneous values in a particular modality can pose a threat to the performance of the system.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

B. Decision-Level Fusion

Decision-level fusion involves the integration of the outputs from independent modality-specific classifiers instead of performing raw feature-level combination. Hence, each modality goes through a separate processing; its associated classifier predicts either a stress or a non-stress label. Such predictions can be combined using approaches like majority voting, weighted averaging, or rule-based logic [16], [17].

However, the decision level provides more flexibility than feature-level fusion. There is no requirement for strict synchronization and uniform representation of features. This condition makes it suitable for real-world applications where modalities might be unavailable due to sensor failure, noise, or user constraints. Such studies have reported comparable, or better robustness by decision-level fusion as compared to feature-level fusion, especially under partial modality availability conditions [14], [16].

C. Hybrid Fusion Approaches

Hybrid fusion techniques are used to incorporate the advantages of both feature-level and decision-level fusion techniques. Based on the hybrid design, certain features may be fused at the feature level, whereas others may be fused at the decision level. Even though hybrid designs offer enhanced performance, they incorporate increased complexity in terms of design and computations [25]. That is why their usage in real-time and wearable stress detection is restricted.

D. Architecture of a Typical Multimodal Stress Detection System

To better understand how fusion strategies are implemented, Fig. 1 illustrates a generic architecture of a multimodal stress detection framework.

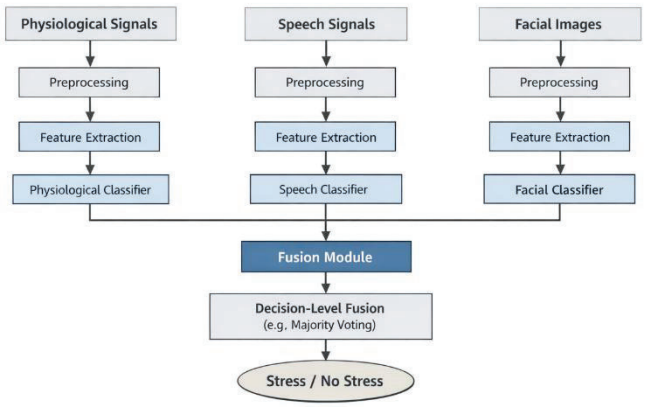

Fig. 1. Generic Architecture of a Multimodal Stress Detection System

Fig 1 marks the modularity in the design of multimodal stress detection systems. As depicted in Figure 1, the system is capable of functioning even if one or more inputs are absent or unreliable. Decision-level fusion is chosen over other fusion techniques because of its improved robustness in stress detection applications [16], [18].

E. Comparative Analysis of Fusion Strategies

A comparison of commonly used fusion strategies in stress detection is provided in Table III, placed below to maintain logical continuity.

Table III. Comparison of Fusion Strategies in Multimodal Stress Detection

Fusion Strategy	Key Characteristics	Advantages	Limitations
Feature-level	Early integration of features	Captures cross-modal correlations	Requires synchronization
Decision-level	Late integration of predictions	Robust to missing modalities	Limited inter-modal learning
Hybrid	Mixed fusion levels	Improved flexibility	High complexity

Table III clearly highlights that the best trade-off between robustness and feasibility is achieved by decision-level fusion. This is exactly why it has been observed that, starting from recent multimodal works given in the literature in [14], [16], fusion has been employed there.

F. Discussion

The choice of the fusion method for decision-making in these applications needs to be made with regard to the requirements, availability, and computation complexity. Although the feature-level fusion technique might be appropriate in a lab setting, decision-level fusion seems suitable for actual applications in stress recognition using heterogeneous sensors in a dynamic environment. The increased reliance on decision-level fusion indicates a drift towards its usability for effective stress detection solutions.

VI. CHALLENGES AND OPEN RESEARCH ISSUES

Although remarkable progress has been made in stress detection research, many technical, practical, and ethical issues still need to be resolved. These issues currently hamper the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

widespread implementation of stress detection system applications. In this section, we will point out the main open issues emerging from a critical analysis of already existing literature.

A. Dataset Limitations and Lack of Standardization

One of the most significant challenges faced while researching in stress detection is that there is no standardized dataset. Nearly all existing datasets have been measured with controlled lab setups with a predefined set of tasks that are used to trigger stress [3], [14]. Although these work well for a benchmarking system, they don't really measure real-world scenarios.

In addition, the data presented in the literature vary greatly with regard to sensor layouts, Lab Annotation task processes, data subjects, and stress label assignment techniques [18], [22]. Thus, evaluation and model generalization are difficult to achieve in a fair manner among several investigations. Imbalanced data, especially in facial stress data, affects model performance and results in biased assessments.

B. Cross-Subject Generalization

Stress responses are very different from each other because of individual physiology, personality characteristics, culture, and levels of coping. Most related studies so far train and test models on splits depending on subjects, which gives rise to overly optimistic performance estimates [4], [21].

Cross-subject stress detection is still an open problem, since models developed on one set of subjects do not generalize well to new subjects. A remedy to this problem requires more data and the need to find new methods of modeling, independent of the subjects.

C. Real-Time and Resource Constraints

In order for stress detection systems to be beneficial, they should be able to run in a real-time manner, taking up less computational time. This is due to the fact that existing systems, using deep models, are computationally intensive and inefficient for execution on wearable devices. This is supported by reference [26].

Further, the concurrent acquisition of data from several sensors consumes more energy due to the continuous operation of the sensor circuitry and the processing of the signals to obtain stress-related information. Energy efficiency while ensuring accuracy remains a major research challenge in the area of stress detection and real-time monitoring systems.

D. Privacy and Ethical Concerns

Stress detection systems are capable of handling extremely personal data in the form of physiological signals, facial images, and audio signals. Such data generates a plethora of concerns with regards to data privacy [22], [25].

Secure data storage, anonymity, and utilization of stress information are of prime importance for user acceptance. Also, transparency in model decision-making is needed for trust creation, especially in applications like healthcare and work monitoring.

E. Robustness to Missing or Noisy Modalities

In real-life situations, sensor data could be incomplete, noisy, or unavailable because of sensor failure, environmental interference, or human actions. Most current multimodal systems model the assumption of having access to all modalities, which is not the case in real-life scenarios as reported in [16].

While decision-level fusion is more robust, still, there is a need to further develop adaptive systems having the ability to cope with variability in modality availability without losing robustness.

F. Summary of Challenges

To provide a concise overview, Table IV summarizes the major challenges and their implications.

Table IV. Key Challenges in Multimodal Stress Detection

Challenge	**Description**	**Impact on System Performance**
Dataset variability	Non-standard datasets	Poor generalization
Class imbalance	Uneven stress labels	Biased accuracy
Subject dependency	Individual stress patterns	Limited scalability

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Challenge	Description	Impact on System Performance
Computational cost	Deep models	Real-time infeasibility
Privacy concerns	Sensitive data	User trust issues

Table IV highlights that most challenges are interconnected and must be addressed jointly rather than in isolation. For instance, improving dataset diversity can enhance generalization while also reducing bias.

G. Research Directions Emerging from Challenges

The problems identified in this section highlight why there is an urgent requirement for effective and adaptive stress detection systems that are ethical as well. Future work needs to be concentrated in creating multimodal datasets, subject-independent models, light models that are portable and suitable for edge processing, and privacy-preserving learning methods.

VII. FUTURE RESEARCH DIRECTIONS

The rapid development in the evolution of sensing technologies, along with machine learning techniques, brings a number of opportunities for stress detection research. Although existing multimodal systems have shown improved performance in contrast to unimodal approaches, there are still several promising directions that remain largely unexplored and thus invite further investigation.

One of the key future directions is the collection of large-scale, real-world multimodal stress datasets. Most current datasets are collected in a controlled laboratory environment and lack ecological validity. The future datasets should reflect natural settings, such as the workplace, academic environment, and daily life scenarios, and should ensure age, gender, cultural, and lifestyle diversity. These kinds of data-sets will thereby be of great use in enhancing the generality of the models.

Another area that is of utmost importance in research entails subject-independent and personalized stress modeling. The hybrid methods, which use a combination of population-level models and personalized adaptation schemes, could offer a compromise between generalization and subject sensitivity. Methods such as transfer learning, meta-learning, and domain adaptation can prove to be the driving force in addressing inter-subject variability.

Light and interpretable models of machine learning must be integrated in the real-time tracking of stress as well. Although models based on deep learning work well, their degrees of complexity and interpretability limit their applicability in the development of wearables. The community must therefore work towards developing efficient models and integrating interpretability in the healthcare setting.

A new area that is also being developed is that of privacy-preserving stress detection. Federated learning, in-device processing, and data anonymization are some methods which can serve to address the concern for privacy as well as make it easier to track stress levels. Ethics in handling stress data must also be factored.

Finally, future stress detection systems need to incorporate more concepts than simply classification and consider other notions like the level of stress or the assessment of stress in the context. Context information such as activity, location, or social interaction can help in gaining a better understanding of the dynamics of stress.

VIII. CONCLUSION

This review paper presented a comprehensive analysis of stress detection techniques, with particular focus on multimodal approaches that integrate physiological signals, speech, and facial expressions. Using a structured taxonomy and a systematic comparison, this paper underlined the limitations inherent in unimodal systems and stressed the advantages of multimodal frameworks concerning robustness, reliability, and real-world applicability.

The review has discussed commonly used datasets, feature extraction techniques, learning models, and fusion strategies. It seemed that decision-level fusion was the most practical and flexible in real-world stress monitoring. Main challenges like dataset variability, subject dependency, computer constraints, and privacy concerns were discussed critically, with some emerging research directions toward their solution.

In essence, this review symbolizes the rise in value across health care, wellness, and human-computer interaction of multimodal detection of stressors. The paper tries to be a useful reference to both researchers and practitioners working toward developing scalable, ethical, and user-centric systems for monitoring stressors by consolidating prior knowledge and presenting open problems.

REFERENCES

[1] H. Selye, The Stress of Life. New York, NY, USA: McGraw-Hill, 1976.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

[2] S. Cohen, J. E. Schwartz, E. Epel, C. Kirschbaum, S. Sidney, and T. Seeman, "Socioeconomic status, race, and diurnal cortisol decline in young adults," Psychosomatic Medicine, vol. 68, no. 1, pp. 41–50, 2006, doi: 10.1097/01.psy.0000195967.51768.ea.

[3] Y. Wang, J. Yang, and X. Chen, "Stress detection and evaluation: A review of wearable sensor systems," IEEE Reviews in Biomedical Engineering, vol. 14, pp. 229–247, 2021, doi: 10.1109/RBME.2021.3105567.

[4] A. Singh and R. Gupta, "Heart rate variability–based stress classification using machine learning," Biomedical Signal Processing and Control, vol. 68, Art. no. 102657, 2021, doi: 10.1016/j.bspc.2021.102657.

[5] J. Kim and S. Park, "Speech-based stress detection using acoustic features and neural networks," Speech Communication, vol. 120, pp. 30–41, 2020, doi: 10.1016/j.specom.2020.02.005.

[6] H. Huang, Z. Zhao, and C. Liu, "Facial expression analysis for stress recognition using deep learning," Pattern Recognition Letters, vol. 124, pp. 45–52, 2019, doi: 10.1016/j.patrec.2019.04.013.

[7] M. Poria, E. Cambria, D. Hazarika, and N. Majumder, "Multimodal affective computing: A survey," Information Fusion, vol. 55, pp. 88–100, 2020, doi: 10.1016/j.inffus.2019.11.008.

[8] J. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 2, pp. 566–576, 2019, doi: 10.1109/TITS.2018.2868287.

[9] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Continuous stress detection using a wrist device," IEEE Access, vol. 8, pp. 48792–48801, 2020, doi: 10.1109/ACCESS.2020.2976665.

[10] Z. Zhang, J. Han, E. Coutinho, and B. Schuller, "Deep learning for emotion recognition in speech," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 908–919, 2020, doi: 10.1109/TASLP.2020.2984228.

[11] D. Neumann and J. Vu, "Robust speech-based stress recognition under noisy conditions," Speech Communication, vol. 129, pp. 1–12, 2021, doi: 10.1016/j.specom.2020.10.004.

[12] R. Roy, T. Bhattacharya, and S. Saha, "Facial stress recognition using convolutional neural networks," IEEE Transactions on Affective Computing, vol. 13, no. 1, pp. 123–134, 2022, doi: 10.1109/TAFFC.2021.3074321.

[13] S. Das, A. Dey, and A. Pal, "Multimodal emotion and stress recognition systems: A review," Applied Soft Computing, vol. 112, Art. no. 107761, 2022, doi: 10.1016/j.asoc.2021.107761.

[14] X. Chen, J. Hernandez, and R. W. Picard, "Wearable sensor-based stress monitoring," Sensors, vol. 19, no. 12, Art. no. 2738, 2019, doi: 10.3390/s19122738.

[15] G. Rigas, C. Katsis, and D. I. Fotiadis, "Stress detection from physiological signals using neural networks," IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 1–14, 2015, doi: 10.1109/TAFFC.2014.2337351.

[16] S. Li, Y. Zhang, and W. Zheng, "Decision-level fusion for multimodal affect recognition," IEEE Access, vol. 9, pp. 13521–13533, 2021, doi: 10.1109/ACCESS.2021.3050097.

[17] J. Lee, S. Kim, and Y. Lee, "Deep learning-based multimodal emotion recognition," Neural Computing and Applications, vol. 32, pp. 10697–10709, 2020, doi: 10.1007/s00521-019-04158-3.

[18] K. Hasan, M. A. Kabir, and A. Rahman, "Physiological signal processing for stress detection: A survey," Sensors, vol. 21, no. 14, Art. no. 4963, 2021, doi: 10.3390/s21144963.

[19] D. Rojas, P. Cano, and J. Ruiz, "CNN-based facial stress detection with imbalanced data," Pattern Recognition, vol. 131, Art. no. 108959, 2022, doi: 10.1016/j.patcog.2022.108959.

[20] A. Kumar, R. K. Tripathi, and S. Jain, "Wearable stress monitoring systems: A review," IEEE Sensors Journal, vol. 22, no. 4, pp. 3211–3220, 2022, doi: 10.1109/JSEN.2021.3133289.

[21] J. A. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161–1178, 1980, doi: 10.1037/h0077714.

[22] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review," IEEE Transactions on Affective Computing, vol. 1, no. 1, pp. 18–37, 2010, doi: 10.1109/T-AFFC.2010.1.

[23] S. Scherer, J. Gratch, and L.-P. Morency, "Multimodal emotion recognition from speech and facial expressions," IEEE Transactions on Affective Computing, vol. 9, no. 4, pp. 410–425, 2018, doi: 10.1109/TAFFC.2017.2714579.

[24] P. Ekman, "An argument for basic emotions," Cognition & Emotion, vol. 6, no. 3–4, pp. 169–200, 1992, doi: 10.1080/02699939208411068.

[25] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey," IEEE Transactions on Pattern

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, 2019, doi: 10.1109/TPAMI.2018.2798607.

[26] Y. Zhao, L. Wang, and X. Liu, "Multimodal stress recognition using deep learning," Computers in Biology and Medicine, vol. 138, Art. no. 104221, 2021, doi: 10.1016/j.compbiomed.2021.104221.

[27] A. 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, doi: 10.1109/T-AFFC.2011.15.

[28] M. Soleymani et al., "A multimodal database for affect recognition and implicit tagging," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 42–55, 2012, doi: 10.1109/T-AFFC.2011.25.

[29] J. Hernandez, P. Paredes, A. Roseway, and M. Czerwinski, "Under pressure: Sensing stress of computer users," CHI Conference Proceedings, pp. 51–60, 2014, doi: 10.1145/2556288.2557165.

[30] L. Shu et al., "A review of emotion recognition using physiological signals," Sensors, vol. 18, no. 7, Art. no. 2074, 2018, doi: 10.3390/s18072074.

[31] S. Zhang and W. Zheng, "Physiological signal-based stress detection," Journal of Biomedical Informatics, vol. 95, Art. no. 103188, 2019, doi: 10.1016/j.jbi.2019.103188.

[32] M. Soleymani and E. Pantic, "Multimodal emotion recognition," IEEE Signal Processing Magazine, vol. 31, no. 1, pp. 128–134, 2014, doi: 10.1109/MSP.2013.2295495.

[33] N. Majumder et al., "Multimodal sentiment analysis using hierarchical fusion," IEEE Intelligent Systems, vol. 34, no. 5, pp. 4–12, 2019, doi: 10.1109/MIS.2019.2929477.

[34] A. Dey et al., "Context-aware affect sensing," ACM Transactions on Interactive Intelligent Systems, vol. 8, no. 2, 2018, doi: 10.1145/3185045.

[35] J. A. Healey, "Wearable and automotive systems for affect recognition," Ph.D. dissertation, MIT, Cambridge, MA, USA, 2000.

[36] M. M. Rahman et al., "Real-time stress monitoring using wearable sensors," IEEE Internet of Things Journal, vol. 7, no. 7, pp. 6006–6016, 2020, doi: 10.1109/JIOT.2020.2970602.

[37] S. Alghowinem et al., "Detecting depression from facial expressions," IEEE Transactions on Affective Computing, vol. 6, no. 3, pp. 227–238, 2015, doi: 10.1109/TAFFC.2014.2347214.

[38] A. Savran et al., "Emotion detection in the wild," IEEE Signal Processing Magazine, vol. 30, no. 2, pp. 32–43, 2013, doi: 10.1109/MSP.2012.2235220.

[39] M. Soleymani, S. Asghari-Esfeden, and Y. Fu, "Analysis of EEG signals for affect recognition," IEEE Transactions on Affective Computing, vol. 7, no. 1, pp. 17–28, 2016, doi: 10.1109/TAFFC.2015.2432736.

[40] Y. Chen and X. Zhang, "Privacy-preserving affective computing," IEEE Access, vol. 9, pp. 113456–113468, 2021, doi: 10.1109/ACCESS.2021.3108294.