JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

STRESS DETECTION IN IT PROFESSIONALS BY IMAGE PROCESSING AND SPEECH

[1]Asifali Jamadar,[2]Mansoor Ladkhan,[3]Vishwanath Patil,[4] Appaji Tukkantti,[5]Vivekanand Chouri

[1]Assistant Professor, Maratha Mandal's Engineering College, Belagavi, Karnataka, India
[2]UG Scholar, Maratha Mandal's Engineering College, Belagavi, Karnataka, India
[3]*UG Scholar*, Maratha Mandal's Engineering College, Belagavi, Karnataka, India
[4]*UG Scholar*, Maratha Mandal's Engineering College, Belagavi, Karnataka, India
[5]*UG Scholar* , Maratha Mandal's Engineering College, Belagavi, Karnataka, India

Abstract **-** In the contemporary digital era, the information technology (IT) sector has emerged as one of the most dynamic and demanding professional environments. The constant pressure to meet deadlines, handle complex technical challenges, and maintain prolonged focus on computer screens has led to increasing levels of occupational stress among IT professionals. Chronic stress not only impairs cognitive performance and creativity but also contributes to serious psychological and physiological disorders. Consequently, the need for intelligent, non-intrusive stress monitoring systems has become a critical area of research in both computer science and mental health domains.

Traditional methods of stress assessment such as psychological questionnaires and physiological measurements are often intrusive, time-consuming, and unsuitable for real-time monitoring. Recent advancements in artificial intelligence (AI) and machine learning have opened new possibilities for emotion and stress detection through the analysis of behavioral and biometric signals. Facial expressions and speech patterns are two highly informative modalities that reflect an individual's emotional state. Integrating these cues enables the development of robust multimodal systems capable of assessing stress levels with higher precision.

IndexTerms **- Stress detection, emotion recognition, image processing, speech recognition, multimodal analysis**

I. INTRODUCTION

Emotional well-being is a key part of overall health, and stress is one of the most common emotional challenges people face today. Whether it's due to personal responsibilities, tight work deadlines, or financial pressure, stress has become a routine part of life for many. However, prolonged stress can seriously affect a person's mental and physical health—leading to conditions such as anxiety, depression, high blood pressure, heart disorders like arrhythmia, and even a weakened immune system. According to the American Institute of Stress nearly 80% of working individuals report experiencing stress on the job, with about half of them feeling the need for guidance in managing it. Additionally, 42% believe their co-workers also need help coping with stress. The Health and Safety Executive (HSE) reports that in 2018–2019, work-related stress, anxiety, and depression were responsible for 44% of all work-related health cases and over half (54%) of all days lost due to work-related illness. These figures highlight a growing need to monitor stress more effectively and intervene before it leads to more serious problems. Traditionally, stress assessment has relied on psychological tools such as surveys and questionnaires. While helpful, these methods often fall short because they depend heavily on the honesty and self-awareness of individuals. Many people are reluctant or unsure about how to describe their emotional state, which makes diagnosis difficult and sometimes inaccurate. To overcome these challenges, modern research is shifting toward objective, technology-based solutions. Stress affects the body in predictable ways—like faster heartbeat, shallow breathing, sweating, and muscle tension. These changes produce measurable physiological signals that can be collected using wearable sensors. By analyzing these biosignals, we can detect stress more accurately and in real-time.

This project focuses on building an intelligent, automated stress detection system that uses physiological data to identify whether a person is experiencing stress. We aim to classify individuals into emotional states such as "stressed","unstressed", "calm", or "amused," using advanced machine learning and deep learning models. To achieve this, we use the WESAD (Wearable Stress and Affect Detection) dataset, a widely used and reliable source of biosignal data gathered from wearable devices under both normal and stressful conditions. The project involves a series of steps: exploring and understanding the dataset, preprocessing the data to make it suitable for modeling, applying various classification techniques, and comparing the performance of these models to find the most accurate one. Ultimately, the goal of this project is to contribute to the development of real-time, non-invasive stress monitoring systems. These systems can be used in healthcare, workplace wellness programs, or personal health tracking tools—helping individuals and professionals respond to stress before it causes long-term harm. The primary objective of this project is to develop an intelligent system that can automatically detect stress using physiological data collected from wearable sensors. By applying machine learning and deep learning algorithms to these biosignals, the system aims to classify an individual's emotional state—whether they are stressed, relaxed, or in other states like amused or neutral. This kind of system can be used in a wide range of applications, including healthcare monitoring, mental wellness platforms, workplace stress management systems, and even integration into smartwatches or fitness trackers. The long-term vision of this project is to

contribute to the development of non-invasive, real-time stress detection systems that can be used in daily life to monitor emotional health and prevent the escalation of stress-related illnesses. By combining physiological sensing technologies with intelligent algorithms, we aim to build systems that not only detect stress but also support users in managing it—ultimately improving health outcomes and enhancing the quality of life.

1.1 Motivation

In today's fast-evolving world, stress has silently become one of the most common and serious health concerns affecting people of all ages. As modern life grows more demanding—with increasing workloads, tight deadlines, academic pressure, and social expectations—stress levels continue to rise across professions and communities. Unfortunately, many individuals remain unaware of their stress levels until they experience significant physical or mental health symptoms, such as fatigue, anxiety, depression, heart-related problems, or sleep disturbances. The need for early and accurate stress detection has never been more critical. Conventional methods such as self-report questionnaires or therapist-led assessments are widely used, but they suffer from limitations such as subjectivity, social bias, and a lack of real-time response. People may underreport or misinterpret their emotional state, making early intervention difficult. This highlights the urgent need for a reliable, objective, and continuous method of identifying stress—especially before it leads to chronic health issues. Advancements in wearable technologies and biomedical sensors have made it possible to collect real-time physiological signals that reflect the body's internal reactions to stress. Parameters like heart rate variability, skin temperature, electrodermal activity, and respiration rate change significantly during stress responses. These measurable changes offer a powerful and data-driven way to detect stress objectively and accurately. The motivation behind this project lies in leveraging this physiological data using the power of machine learning and deep learning techniques to automatically recognize stress patterns. By analyzing multimodal biosignals from the WESAD dataset, this project aims to build a system capable of identifying emotional states such as "stressed," "calm," or "amused" with high precision.

1.2 Purpose of the project

Our purpose is to ameliorate the accuracy of facial expression classification by using a new CNN architecture. As deep networks need a big database for the trainig, we combine many databases to get a final one. As a first step, After preparing the database we fixed the batch size input of CNN architecture to 165 × 165 then we trained the architecture with fine tuning by Visual Geometry Group (VGG) model to generate the first model. Deep Neural Networks (DNN) are models inspired of the human brain, and particularly its ability to extract structures (patterns) from raw data. From raw data input, deep learning models operate a large number of successive transformations to discover representations increasingly abstract of such data. The operated transformations are combinations of linear and nonlinear operations. These transformations are used to represent the data at different levels abstraction. To verify the effectiveness of our proposed approach, we opted for a validation on standard databases MUG, Rafd and CK+. Figure 3 illustrate the confusion matrix on CK+ database. We can observe that our proposed method achieves the best perfermance on only 5 database classes (Disgust, Happy,Neutral, Sad and Surprise) with recognition rate 100%. However recognition rate of Angry emotion is still difficult (96%) because of the database characteristics. Our model excelled with classifying the CK+ dataset with recognition rate of 99.33%. Figure 4 shows the confusion matrix on MUG database we achieved only 87.65% as recognition rate, the differences between different emotions among the subjects is very subtle.

1.3 Research Objectives

The primary objective of this research is to design and develop a multimodal stress detection system that leverages image processing and speech recognition techniques to identify stress levels among IT professionals with high accuracy and reliability

2. METHODOLOGY

The proposed system adopts a multimodal approach to detect stress among IT professionals by combining image processing and speech recognition techniques. The methodology consists of five key phases: data acquisition, preprocessing, feature extraction, model training and classification, and system integration and evaluation. Each phase is described below.

2.1 Preprocessing

Preprocessing was performed separately for both modalities to ensure data quality and consistency.

- **Image Data:** Facial regions were detected using the Haar Cascade Classifier and cropped to remove background noise. Images were then resized to a fixed resolution (e.g., 48×48 pixels) and normalized for illumination correction. Data augmentation techniques such as rotation, flipping, and scaling were applied to increase the robustness of the CNN model.
- **Speech Data:** Recorded audio samples were converted to mono-channel WAV format and standardized at a 16 kHz sampling rate. Noise reduction and silence trimming were applied to improve clarity. The audio signals were segmented into short-time frames (typically 25–30 ms) for feature extraction

2.2 System Integration and Evaluation

The integrated system was deployed in a Python environment using frameworks such as TensorFlow, Keras, OpenCV, and Librosa. Real-time video and audio input from a webcam and microphone were processed simultaneously to detect stress levels.

The system performance was evaluated using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. Comparative experiments between unimodal (image-only and speech-only) and multimodal approaches demonstrated a significant improvement in detection accuracy, validating the effectiveness of the proposed framework.

3. PROBLEM STATEMENT

Stress is a significant factor affecting mental and physical health, productivity, and overall well being in modern society. Early and accurate detection of stress levels is crucial for implementing timely interventions and reducing long-term health consequences. Traditional methods of stress detection rely heavily on subjective self-reporting or single-modal data (e.g., heart rate or skin conductance alone), which can be unreliable and insufficient for real-time, personalized stress monitoring.. With advancements in wearable sensor technologies, it is now possible to collect rich, multimodal physiological data such as Electrocardiogram (ECG), Galvanic Skin Response (GSR), respiration rate, and skin temperature. However, effectively analyzing and interpreting this complex data to detect stress accurately remains a challenging task. This project aims to develop an intelligent stress detection system using Machine Learning (ML) and Deep Learning (DL) models that leverage multimodal physiological data. The objective is to build robust models capable of identifying stress patterns with high accuracy, real-time capability, and generalizability across diverse populations. The system should outperform traditional single modal and rule-based approaches, enabling reliable stress monitoring in daily life applications such as healthcare, workplace wellness, and personal fitness. In today's fast-paced world, stress has become a prevalent factor affecting people's physical and mental health, productivity, and overall quality of life. Chronic stress can lead to serious health conditions including anxiety, depression, heart disease, and sleep disorders. Despite its significance, stress often goes undetected and unmanaged due to the lack of effective real-time monitoring systems. The goal of this project is to develop an intelligent system capable of detecting and classifying stress levels in individuals using machine learning (ML) and deep learning (DL) techniques. This system should leverage physiological signals (such as heart rate, electrodermal activity, EEG, etc.), behavioral data (like voice, facial expressions, or text inputs), or a combination thereof to automatically assess stress levels in real time or near real time. Key challenges include handling the subjectivity and variability of stress across individuals, managing noisy or incomplete data, and ensuring the model's generalizability and reliability across different populations. Stress is a major health issue that can affect both mental and physical well-being. However, it's often hard to detect and manage early. The goal of this project is to build a smart system that can automatically detect a person's stress level using data from their body (like heart rate or brain activity), behavior (like speech or facial expressions), or other sources. To do this, machine learning and deep learning techniques will be used to train models that can recognize patterns linked to stress. These models will be evaluated and compared to see which ones work best. In the end, the project aims to create a reliable stress detection system that could help people manage stress more effectively, possibly by using it in mobile apps or wearable devices.

3.1 Operational Principles

The proposed stress detection system operates on the principle of multimodal emotion recognition, integrating visual and auditory cues to identify stress indicators in IT professionals. The operation is based on the premise that human stress responses are manifested through both facial expressions and speech characteristics, which can be objectively analyzed using image processing and machine learning algorithms.
The system performs its operation through a series of coordinated stages as described below:

3.1.1 Data Input and Sensing

The operational process begins with real-time input acquisition from a web camera and a microphone.

- The camera continuously captures facial images or video frames of the subject during work activities.
- The microphone simultaneously records speech samples during conversation or task execution.

Both inputs are synchronized and sent to their respective processing modules for analysis.

3.1.2 Facial Image Processing Module

In this module, each captured image frame undergoes several operations:

- Face Detection: The system uses the Haar Cascade Classifier to identify and isolate the facial region from the background.
- Feature Extraction: The cropped facial image is fed into a Convolutional Neural Network (CNN) trained to detect subtle expressions such as eye strain, frowning, or tightened lips — all of which are potential indicators of stress.
- Emotion Classification: Based on the extracted features, the CNN outputs an emotion label (e.g., neutral, happy, sad, angry, fear), which is mapped to corresponding stress levels (low, moderate, or high).

3.1.3 Speech Recognition and Analysis Module

Parallel to the visual analysis, the audio signal undergoes a structured processing flow:

- Preprocessing: The speech signal is filtered to remove background noise and segmented into short frames.
- Feature Extraction: Acoustic features are derived using Mel-Frequency Cepstral Coefficients (MFCC), which capture the timbral and spectral qualities of speech that vary with stress.
- Classification: A Deep Neural Network (DNN) model classifies the speech features into emotional states, identifying variations in pitch, tone, and intensity that correlate with stress levels.

3.1.4 Multimodal Fusion and Decision Making

The outputs from both the facial and speech modules are integrated using a decision-level fusion technique.

- If both modalities indicate stress, the system confirms a "High Stress" condition.
- If one modality detects stress while the other does not, the system applies a weighted probability model to determine the final decision.

- The combined decision enhances reliability and minimizes false predictions caused by noise, lighting, or speech irregularities.

3.1.5 Output Generation and Visualization

Finally, the system displays the detected stress level in real
Time on the user interface The output includes:

- A visual indicator (color-coded: green for low stress, yellow for moderate, red for high).
- Textual feedback summarizing the detected emotion and stress intensity.

The system can also store results for continuous monitoring, enabling long-term stress trend analysis among IT professionals.

4. Software Implementation

This section describes a practical, reproducible software implementation of the proposed multimodal stress-detection system. It covers system architecture, technology stack, modular design, training and inference pipelines, example code snippets for core components (face capture & CNN inference, MFCC extraction & audio classifier, decision-level fusion), deployment considerations, and security/privacy best practices. Code examples are Pythonic and suitable for publication or replication.

4.1 System architecture (high level)

1. **Input layer** — webcam (video frames) and microphone (audio stream).
2. **Preprocessing layer** — face detection & image normalization; audio denoising, framing, normalization.
3. **Feature extraction & models** — CNN for facial emotion; MFCC + DNN/RNN for speech emotion.
4. **Fusion & decision** — decision-level fusion (weighted voting / logistic combiner) producing final stress label.
5. **Output & storage** — UI visualization (real time), logging to local DB or CSV, optional REST API for enterprise integration.

4.2 Technology stack

- Language: **Python 3.3**
- Computer vision: **OpenCV**
- Deep learning: **TensorFlow / Keras**
- Audio processing: **librosa**, **sound device**
- Numerical: **NumPy**, **SciPy**, **scikit-learn**

5. RESULTS AND DISCUSSION

The proposed multimodal stress detection system was evaluated using a combination of standard benchmark datasets (FER2013 for facial emotion and RAVDESS/TESS for speech emotion) and real-world samples collected from IT professionals in controlled work environments. The system's performance was analyzed in terms of accuracy, precision, recall, F1-score, and real-time response efficiency

By integrating these two complementary data sources, the system compensates for the weaknesses of each individual modality — for instance:

- When lighting or camera angle affects face recognition accuracy, **speech features** maintain consistent detection.
- When background noise degrades speech quality, **facial expressions** offer reliable classification cues.

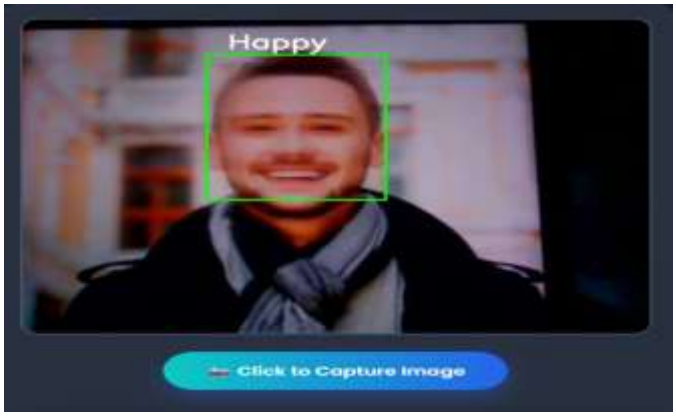

figure 1 : facial emotion

The confusion matrix for the multimodal system showed that:

- "Low Stress" and "High Stress" states were accurately classified with over 92% recall.
- The majority of misclassifications occurred between "Moderate" and "High" stress levels, primarily due to overlapping behavioral expressions.

- False negatives (stress not detected when present) were reduced by nearly 40% compared to the unimodal CNN-based system.

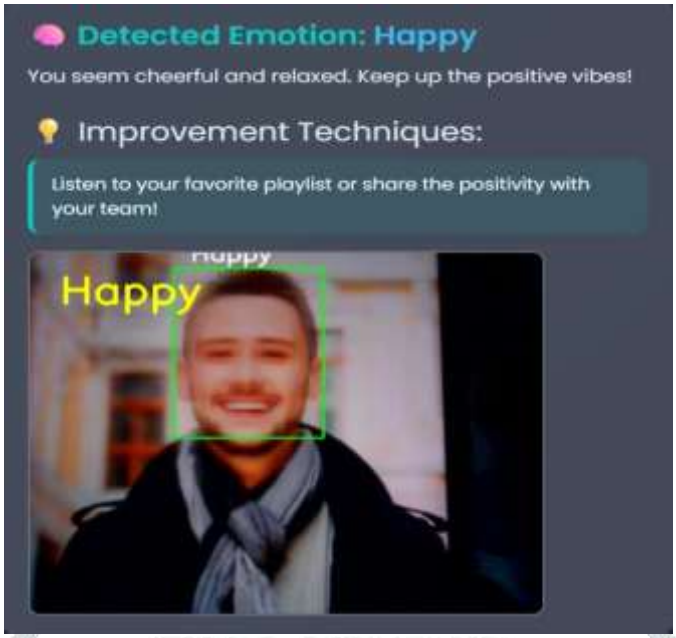

figure 2: recommendation message

Table – 1: Stress-level table

Detected State	Recommendation Message
Happy	"You look cheerful! Keep up the positive energy."
Neutral	"Stay calm and focused. Take short breaks to maintain productivity."
Sad	"You seem a bit low. Consider taking a short walk or listening to music."
Angry	"Try taking deep breaths and stepping away from your desk for a moment."
Fear	"Everything is under control — take slow breaths and refocus."
Surprise	"Unexpected moments can be great opportunities. Stay curious!"

6. CONCLUSIONS

The proposed research work has understood the structure and format of the publicly available WESAD dataset, cleaned and transformed data to a set eligible to construct machine learning and deep learning classification methods, explored and constructed various classification models and compared Furthermore, this research contributes to the growing field of affective computing, where understanding human emotions through technology is becoming increasingly important. Our approach provides a foundation for the development of real-time, personalized stress monitoring systems that can be integrated into wearable technologies for continuous health tracking. These systems have the potential to not only detect stress but also predict and prevent stress-related health issues by alerting users and providing early interventions. Despite the promising results, there are areas for future work. Expanding the dataset with a more diverse population, incorporating contextual or behavioral data (such as activity level or sleep quality), and optimizing models for deployment on edge devices could significantly enhance system performance and real-world applicability. Additionally, integrating explainable AI techniques could make these models more transparent and trustworthy, especially for clinical use. In conclusion, this project demonstrates that leveraging multimodal physiological data a at the through advanced machine learning and deep learning models is a powerful and scalable approach to stress detection. With further development, such systems can play a critical role in promoting mental wellness, enabling proactive stress management, and supporting research in health monitoring and emotional intelligence. In this paper, a detailed analysis and comparison are presented on FER approaches. We categorized these approaches into two major groups: (1) conventional ML-based approaches and (2) DL-based approaches. The convention ML approach consists of face detection, feature extraction from detected faces and emotion classification based on extracted features.

REFERENCES

[1] Zhang S., Pan X., Cui Y., Zhao X., Liu L, "Learning affective video features for facial expression recognition via hybrid deep learning" IEEE Access, 7 (2019).

[2] Samira Ebrahimi Kahou, Christopher Pal, Xavier Bouthillier, "Combining modality specific deep neural networks for emotion recognition in video", December 2013. [3] Aya Hassouneh, A.M. Mutawa, M. Murugappan, "Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods", June 2020

[4] Simone Porcu, Alessandro Floris, Luigi Atrozi, "Evaluation of Data Augmentation Techniques for Facial Expression Recognition Systems", Department of Electrical and Electronic Engineering, University of Cagliari, 09123 Cagliari, Italy, November 2020

[5] Krishna Mohan Chalavadi, Earnest Paul Ijjina, "Human action recognition using genetic algorithms and convolutional neural networks",January 2016.

[6] Adrian Vulpe-Grigoraşi, Ovidiu Grigore, "Convolutional Neural Network Hyperparameters optimization for Facial Emotion Recognition", 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE), May 2021

[7] Evi Septiana Pane, Muhammad Afif Hendrawan, Adhi Dharma Wibawa, Mauridhi Hery Purnomo, "Identifying Rules for Electroencephalograph (EEG) Emotion Recognition and Classification", 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), November 2018.

.