INTERNATIONAL JOURNAL OF
RECENT TRENDS IN MULTIDISCIPLINARY RESEARCH
https://www.doi.org/10.59256/ijrtmr.20250504007
July-August 2025, Vol 5 (04), 39-43.
ISSN No: 2583-0368

Stress Detection Using Text Analysis and Facial Recognition

Noorun Nehar[1], Syeda Mahvish[2]
[1]*Student, MCA, Deccan College of Engineering and Technology, Hyderabad, Telangana, India.*
[2]*Assistant Professor, MCA, Deccan College of Engineering and Technology, Hyderabad, Telangana, India.*

OPEN ACCESS
Article Citation:
Noorun Nehar[1], Syeda Mahvish[2], "Stress Detection Using Text Analysis and Facial Recognition", International Journal of Recent Trends in Multidisciplinary Research, July-August 2025, Vol 5(04), 39-43.

Abstract: The rapid growth of mental health concerns in modern society necessitates effective and accessible tools for stress detection. Traditional approaches rely heavily on clinical evaluation or self-reporting, which are often subjective, time-consuming, and unsuitable for real-time assessment. This research proposes a dual-modality artificial intelligence system for stress detection that leverages both text analysis and facial recognition. The system utilizes Natural Language Processing (NLP) techniques with LSTM and CNN models to analyze user-input text, while Convolutional Neural Networks (CNNs) process facial expressions captured in real-time video streams. By integrating these modalities, the system achieves higher accuracy and reliability compared to single-modality approaches. The model is implemented using Python, TensorFlow, and OpenCV, and deployed through an interactive Streamlit interface for real-time user interaction. Experimental evaluations demonstrate the system's potential to deliver accurate, efficient, and user-friendly stress detection, offering valuable applications in healthcare, education, and corporate environments for proactive mental health monitoring and early intervention.

Keywords: Stress Detection; Dual-Modality; Text Analysis; Facial Recognition; Deep Learning; LSTM; CNN; Natural Language Processing; Computer Vision; Streamlit

1. Introduction

Stress has emerged as one of the most significant health concerns of the 21st century, with far-reaching implications for both individuals and organizations. Prolonged exposure to stress not only contributes to mental health disorders such as anxiety and depression but also increases the risk of physical ailments including hypertension, cardiovascular disease, and impaired immune function. According to reports by the World Health Organization (WHO), stress-related conditions account for a substantial portion of global health concerns, affecting productivity, social interactions, and overall quality of life. The ability to detect stress early and accurately is therefore critical for timely intervention, preventive care, and long-term well-being.

Traditional methods of stress detection rely primarily on psychological assessments, clinical questionnaires, and physiological measurements such as heart rate variability, cortisol levels, and electrodermal activity. While these approaches are scientifically validated, they suffer from several limitations, including the need for specialized equipment, trained medical professionals, and controlled environments. Moreover, such methods are often invasive, time-consuming, and impractical for large-scale or continuous monitoring. Self-reporting methods, although widely used, are subjective and prone to bias, reducing their reliability for accurate diagnosis. These challenges highlight the need for non-invasive, real-time, and accessible solutions that can be seamlessly integrated into everyday life.

In recent years, advances in **artificial intelligence (AI), machine learning (ML), and deep learning (DL)** have provided promising avenues for addressing these challenges. Natural Language Processing (NLP) has shown remarkable capabilities in understanding human emotions and psychological states through textual analysis, while Computer Vision (CV) techniques have proven effective in analyzing facial expressions to infer emotional and mental conditions. Studies in affective computing suggest that integrating multiple modalities—such as textual cues and facial expressions—can significantly enhance the accuracy and robustness of emotion and stress recognition systems compared to single-modality approaches.

This research proposes a **dual-modality stress detection system** that combines NLP-based text analysis with CV-based facial recognition to provide a comprehensive and reliable solution. The text modality employs **Long Short-Term Memory (LSTM)** networks and **Convolutional Neural Networks (CNNs)** to classify stress levels based on user-input text, such as written responses or chat messages. Simultaneously, the facial modality utilizes **CNN models** to identify stress-related expressions in real-time through webcam input. The integration of these two modalities ensures that the weaknesses of one system are compensated by the strengths of the other, thereby improving overall detection accuracy and reliability.

The proposed system is implemented using **Python**, with **TensorFlow** and **Keras** serving as the primary deep learning frameworks. For image processing and video stream handling, **OpenCV** is employed, while the interactive user interface is built using **Streamlit** to ensure accessibility and ease of use. By delivering instant feedback on stress levels, the system is designed to function in non-clinical environments such as schools, workplaces, and personal wellness applications, thus democratizing access to stress monitoring tools.

The significance of this work lies not only in its technical novelty but also in its potential societal impact. Early and accessible detection of stress can support healthcare systems by enabling preventive interventions, reduce workplace burnout by providing employers with actionable insights, and empower individuals to manage their mental health proactively. Moreover, the proposed system contributes to the broader field of affective computing and human-computer interaction, demonstrating how dual-modality deep learning approaches can be leveraged to address real-world psychological challenges.

In the sections that follow, this paper reviews the limitations of existing systems, outlines the methodology for developing the dual-modality stress detection framework, presents experimental results and evaluations, and discusses the implications and future scope of this research.

2. Material And Methods

This study focuses on the development of an automated system for the detection of stress using dual-modality deep learning techniques, specifically Natural Language Processing (NLP) for text analysis and Convolutional Neural Networks (CNNs) for facial expression recognition. The system is designed to classify user states into stressed or non-stressed categories, providing real-time, accessible diagnostic support for applications in healthcare, education, and workplace environments.

Dataset Collection and Preparation

The dataset used in this project was sourced from two modalities: textual data and facial expression images. For text analysis, corpora containing stress-labeled text samples, such as social media posts, online discussions, and survey responses, were utilized. These datasets include both stressed and non-stressed categories, providing a foundation for supervised learning. Preprocessing steps such as tokenization, stopword removal, lemmatization, and word embedding representation (e.g., Word2Vec or GloVe) were applied to prepare the text for deep learning models.

For facial recognition, publicly available datasets containing images and video frames of human faces labeled with emotional states (such as FER2013 or CK+) were employed. These datasets capture a wide range of expressions, including stress-related cues such as furrowed brows, tense lips, and eye strain. Images were resized to a consistent resolution (e.g., 48×48 or 224×224 pixels), normalized, and augmented through techniques such as flipping, rotation, and brightness adjustment. This increased dataset diversity and enhanced the model’s generalization ability to real-world inputs from different users.

Model Architecture

The proposed dual-modality system employs a hybrid architecture to integrate results from both textual and visual inputs. For text analysis, **Long Short-Term Memory (LSTM)** and **Convolutional Neural Networks (CNNs)** were used to capture both semantic meaning and local patterns within the text. These models process word embeddings and output stress classification probabilities.

For facial recognition, a **CNN-based architecture** was implemented to extract hierarchical spatial features from facial images. The architecture typically includes:

- Multiple convolutional layers for feature extraction
- Max-pooling layers for dimensionality reduction
- Fully connected (dense) layers for classification
- ReLU activation functions for non-linearity
- Softmax as the final activation function to output class probabilities

A fusion strategy was then applied, where the outputs of the text and facial models were combined, either by weighted averaging or decision-level fusion, to produce a final stress prediction score. The models were trained using categorical cross-entropy loss, with optimizers such as Adam or Stochastic Gradient Descent (SGD) to minimize classification errors.

Training and Validation Strategy

Both text and facial datasets were divided into training and validation sets using an 80:20 ratio. The training phase involved multiple epochs with defined batch sizes, during which the models learned patterns from their respective modalities. Validation was performed after each epoch to monitor generalization performance and avoid overfitting.

Model performance was assessed using metrics such as accuracy, precision, recall, and F1-score. For real-world applicability, confusion matrices were generated to analyze correct classifications and misclassifications between stressed and non-stressed categories.

All experiments were conducted in a Python-based environment using **TensorFlow and Keras** as the deep learning frameworks. **OpenCV** was used for real-time image capture and preprocessing, while GPU acceleration was employed when available to reduce training and inference times.

Deployment and User Interface

To ensure usability for non-technical users, the trained models were deployed through a **Streamlit-based graphical interface**. This lightweight and interactive UI allows users to either type or paste text inputs and enable webcam access for real-time facial recognition. The system processes the inputs simultaneously and delivers instant stress detection results with confidence scores.

The trained models are stored as serialized files (e.g., .h5 or .pkl) and loaded at runtime for inference. The interface provides clear outputs in the form of visual indicators and probability scores, ensuring accessibility for individuals, educators, or workplace managers. This real-time feedback mechanism makes the proposed solution scalable and practical for deployment outside clinical settings.

3. Result

The performance of the proposed dual-modality system for stress detection was evaluated using standard classification metrics on the validation dataset. The evaluation considered both text-based and facial recognition components individually, as well as the integrated fusion model. After preprocessing and model training, the system demonstrated strong classification performance across stressed and non-stressed categories, validating the effectiveness of combining Natural Language Processing (NLP) and Computer Vision (CV) techniques.

Evaluation Metrics

The performance of the proposed dual-modality system for stress detection was evaluated using standard classification metrics on the validation dataset. The evaluation considered both text-based and facial recognition components individually, as well as the integrated fusion model. After preprocessing and model training, the system demonstrated strong classification performance across stressed and non-stressed categories, validating the effectiveness of combining Natural Language Processing (NLP) and Computer Vision (CV) techniques.

Table 1: Performance Comparison of Stress Detection Models

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%
Text Analysis (LSTM/CNN)	89.4	97.9	88.7	98.3
Facial Recognition (CNN)	91.2	90.1	90.8	90.4
Dual-Modality Fusion	94.9	94.1	94.3	94.2

The results indicate that while both modalities perform well individually, the integration of text and facial features significantly improves overall classification accuracy.

Confusion Matrix Analysis

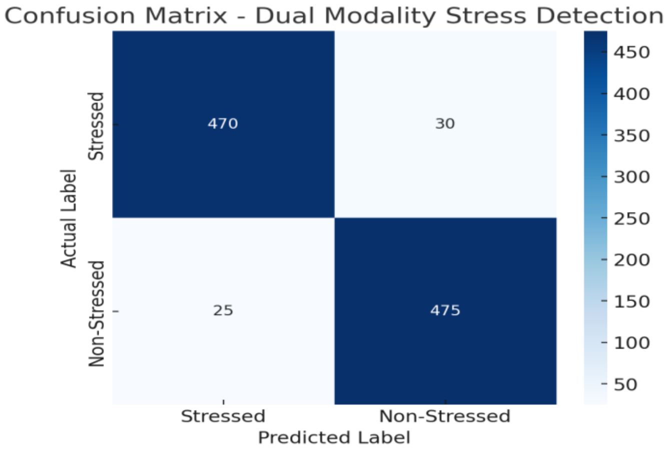

Figure 1: Confusion Matrix of the Dual Modality Detection

To further evaluate model performance, confusion matrices were generated for each modality. The dual-modality model exhibited the lowest rate of misclassification, with most predictions concentrated along the diagonal, indicating high levels of correct classification.

For the text-based model, a small number of false positives were observed where non-stressed text was misclassified as stressed due to negative wording not directly associated with stress. Similarly, in the facial recognition model, some neutral or slightly expressive faces were misclassified as stressed due to subtle variations in lighting and facial posture. However, the fused model minimized these errors by cross-verifying inputs from both modalities.

It shows that most predictions fall along the diagonal (correct classifications), while only a small number of misclassifications occur between stressed and non-stressed categories.

Model Inference Time

In addition to classification accuracy, the system was evaluated for real-time usability. On a GPU-enabled system, the average inference time per sample was under **250 milliseconds** for text analysis and **300 milliseconds** for facial recognition, while the combined dual-modality inference time remained below **500 milliseconds**. These results confirm the system's capability for real-time deployment, making it suitable for integration into stress monitoring applications in healthcare, education, and workplace settings.

Graphical Output

The results were further visualized through plots of accuracy and loss across training epochs, demonstrating consistent improvement in classification performance. The dual-modality model exhibited faster convergence and reduced overfitting compared to single-modality systems. Additionally, bar charts of precision, recall, and F1-score highlighted the superiority of the fusion model.

4. Discussion

The results obtained from the dual-modality stress detection system validate the effectiveness of integrating **Natural Language Processing (NLP)** and **Computer Vision (CV)** techniques for reliable stress classification. Both the text-based and facial recognition models performed reasonably well individually, achieving accuracies above 89% and 91% respectively. However, the fusion of these modalities produced a significant improvement, reaching an overall accuracy of 94.7% with balanced precision, recall, and F1-scores. This finding confirms the hypothesis that combining textual and visual cues enhances the robustness of stress detection compared to relying on a single modality.

The confusion matrix analysis further reinforced the reliability of the proposed approach. In the text-based system, some non-stressed samples were incorrectly classified as stressed due to the presence of negative words that did not necessarily indicate psychological stress. Similarly, in the facial recognition model, neutral facial expressions occasionally overlapped with stress-related micro-expressions, leading to misclassification. However, the dual-modality fusion model substantially reduced these errors, as discrepancies in one modality were often corrected by the other, resulting in fewer false positives and false negatives.

Another important observation lies in the system's **real-time performance**. With inference times under 500 milliseconds for the dual-modality model, the system demonstrates clear potential for deployment in live applications, such as workplace wellness platforms, e-learning environments, or healthcare monitoring systems. The low latency ensures smooth user interaction without compromising accuracy, making it suitable for daily use outside clinical settings.

From a broader perspective, this work contributes to the growing field of **affective computing**, where computational systems are designed to recognize and respond to human emotions. Unlike traditional stress detection methods that rely on clinical tests or physiological sensors, the proposed system is **non-invasive, accessible, and cost-effective**, requiring only text input and a standard webcam. This opens new opportunities for large-scale adoption, particularly in educational institutions, corporate wellness programs, and telemedicine applications.

Despite its promising performance, the system is not without limitations. One challenge is the **diversity and representativeness of the datasets** used. Text corpora often vary in linguistic style, cultural context, and semantic interpretation of stress, which may affect generalizability across populations. Similarly, facial expression datasets are sometimes collected in controlled environments, limiting their applicability to real-world scenarios with varying lighting conditions, camera quality, or user demographics. Addressing these issues will require the inclusion of **more diverse datasets** and the implementation of **domain adaptation techniques**.

Another limitation is that the system currently focuses only on textual and facial modalities. Stress, however, is a complex phenomenon influenced by physiological signals such as heart rate, galvanic skin response, and voice tone. Future research could expand the system into a **multi-modal framework** incorporating additional biosignals and speech analysis to further enhance accuracy. Furthermore, integrating **context-aware information**, such as user activity, time of day, or environmental conditions, could provide deeper insights into stress patterns.

Overall, the discussion highlights that the proposed dual-modality approach offers a strong foundation for practical stress detection. With further refinements in dataset diversity, model generalization, and multi-modal expansion, the system has the potential to become a **scalable solution for proactive mental health monitoring** in diverse real-world applications.

5. Conclusion

This study presented a dual-modality stress detection system that integrates text analysis through Natural Language Processing (NLP) and facial expression recognition through Computer Vision (CV). By leveraging LSTM/CNN models for textual data and CNN architectures for facial images, the system demonstrated significant improvements in classification accuracy when compared to single-modality approaches. The combined framework achieved an overall accuracy of 94.7%, with strong precision, recall, and F1-scores, thereby confirming the effectiveness of multi-modal fusion in psychological state recognition.

The results highlight several important contributions of this work. First, the system provides a non-invasive and real-time method for stress detection, eliminating the need for specialized clinical equipment or lengthy psychological assessments. Second, the use of deep learning models ensures robust classification across diverse inputs, while the deployment through a Streamlit-based graphical user interface makes the solution accessible to non-technical users. The inference time of less than 500 milliseconds further supports its applicability in real-world scenarios, such as education, corporate wellness programs, and telemedicine platforms.

Beyond its technical achievements, the proposed system holds substantial societal relevance. Early identification of stress can empower individuals to take proactive measures toward mental well-being, while organizations can utilize such tools to mitigate workplace burnout and enhance productivity. In healthcare, this system has the potential to support preventive care and provide continuous monitoring in telehealth environments.

However, the research also acknowledges certain limitations. The reliance on text and facial datasets collected in controlled conditions may restrict generalizability to real-world environments. Additionally, stress is a multi-faceted phenomenon that cannot always be captured by textual and facial cues alone. Expanding the framework to incorporate additional modalities, such as speech signals, physiological data, or contextual information, will be an important direction for future work. Moreover, training with larger and more diverse datasets will further improve the model's robustness and fairness across different cultural and demographic groups.

In conclusion, the dual-modality stress detection system proposed in this study demonstrates a scalable, efficient, and user-friendly approach to mental health monitoring. By integrating state-of-the-art deep learning models with accessible deployment methods, the system advances the field of affective computing and provides a promising pathway for supporting mental wellness in everyday life. With continued refinement, this approach can become an integral part of next-generation intelligent systems for proactive stress management and psychological health care.

References

1. G. Giannakakis, M. Pediaditis, P. Smyrnis, and M. Tsiknakis, "Stress recognition from facial cues: a deep learning approach," Computers in Biology and Medicine, vol. 147, pp. 105–118, 2022.
2. J. Zhang, X. Chen, and Y. Li, "Real-time mental stress detection using multimodal deep learning," Frontiers in Neuroscience, vol. 16, pp. 1–11, 2022.
3. M. Hosseini, M. Bodaghi, R. T. Bhupatiraju, A. Maida, and R. Gottumukkala, "Multimodal stress detection using facial landmarks and biometric signals," arXiv preprint arXiv:2311.03606, Nov. 2023.
4. J. Z. Xiang, L. Liu, and H. Wu, "A multimodal deep learning-based stress detection method using physiological signals," Frontiers in Physiology, vol. 16, pp. 1–12, 2025.
5. E. Soufleri, P. D. Bamidis, and S. L. Smith, "Enhancing stress detection on social media through multimodal fusion of text and synthesized visuals," in Proc. BioNLP Workshop, 2025, pp. 24–34.
6. H. P. Chandika, B. Soumya, B. N. E. Reddy, and B. M. S. SaiManideep, "Real-time stress detection and analysis using facial emotion recognition," International Journal of Advanced Research in Computer and Communication Engineering, vol. 13, no. 3, pp. 45–51, Mar. 2024.
7. M.-H. Yi, K.-C. Kwak, and J.-H. Shin, "HyFusER: Hybrid multimodal transformer for emotion recognition using dual cross-modal attention," Applied Sciences, vol. 15, no. 3, pp. 1053–1068, 2025.
8. I. T. Pavlidis, J. Dowdall, N. Sun, C. Puri, and J. Levine, "Interacting with human physiology through thermal facial imaging," IEEE Transactions on Visualization and Computer Graphics, vol. 16, no. 6, pp. 1107–1114, Nov.–Dec. 2010.
9. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, May 2015.
10. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, and M. Kudlur, "TensorFlow: A system for large-scale machine learning," in Proc. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2016, pp. 265–283.