Multimodal Deep Learning Framework for Employee Stress Detection in the Workplace

Srungarapati Pavan
Department of Artificial Intelligence and Data Science
Central University of Andhra Pradesh, Ananthapuramu, India
Email: pavansrungarapati123@gmail.com

Dr. P. Sumalatha
Assistant Professor
Department of Artificial Intelligence and Data Science
Central University of Andhra Pradesh, Ananthapuramu, India
Email: dr.sumalatha@cuap.edu.in

***Abstract*—Workplace stress is a pervasive issue impacting employee well-being and organizational productivity. This study proposes a multimodal AI framework for detecting stress in employees using structured workplace data, such as age, job role, workload, overtime, and support metrics. Machine learning (ML) algorithms, including Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Na¨ıve Bayes, were benchmarked against deep learning (DL) models, such as Convolutional Neural Networks (CNN), Multilayer Perceptron (MLP), Bidirectional Long Short-Term Memory (BiLSTM), and Attention-BiLSTM. Experimental results demonstrate exceptional performance, with Decision Tree and Random Forest achieving accuracies of 99.83% and 99.67%, respectively, and CNN and BiLSTM reaching 99.00% and 98.50%. Explainable AI (XAI) techniques, including SHAP and LIME, were integrated to enhance model interpretability, ensuring actionable insights for HR teams. The framework offers a scalable, ethical, and deployable solution for proactive stress management, with potential for real-time monitoring using physiological and behavioral data.**

KeywordsStress detection, workplace wellness, machine learning, deep learning, multimodal data, explainable AI

I. Introduction

Workplace stress is a critical challenge in modern organizations, contributing to mental and physical health issues such as depression, anxiety, and cardiovascular diseases [1]. Global studies indicate that approximately 25% of employees experience chronic stress, particularly in high-pressure sectors like Information Technology (IT), where long hours, heavy workloads, and tight deadlines are common [2]. Chronic stress leads to reduced productivity, increased absenteeism, and higher turnover rates, costing organizations billions annually [3]. Traditional stress detection methods, such as self-reported surveys and HR interviews, are subjective, time-consuming, and lack scalability, often identifying stress only after it has impacted performance or health [4].

Artificial Intelligence (AI), particularly machine learning (ML) and deep learning (DL), offers a transformative approach to address these limitations. By analyzing structured workplace data (e.g., age, gender, job role, workload, overtime, pressure, support, and sick leaves) and simulated behavioral patterns, AI models can predict stress levels with high accuracy, enabling proactive interventions. This research develops a multimodal AI-driven framework for stress detection, aiming to create a scalable, interpretable, and ethical system for workplace wellness. The specific objectives are:

- To construct a multimodal dataset combining structured workplace data and simulated behavioral patterns.
- To develop and evaluate DL models (CNN, MLP, BiLSTM, Attention-BiLSTM) and compare them with ML algorithms (Decision Tree, Random Forest, SVM, KNN, Na¨ıve Bayes).
- To identify key stress predictors using explainable AI (XAI) techniques.
- To propose a deployable framework for real-time stress monitoring in organizational settings.

This paper is organized as follows: Section II reviews related work, Section III details the methodology, Section IV presents experimental results, Section V discusses findings, and Section VI concludes with future directions.

II. Literature Review

The application of AI in stress detection has gained significant attention, with studies leveraging both ML and DL to analyze diverse data sources. Miranda-Correa et al. [1] introduced the AMIGOS dataset, which includes neurophysiological signals (EEG, ECG, GSR) and high-resolution videos for affect and mood analysis, highlighting the potential of multimodal data. Koldijk et al. [7] developed the SWELL project, using unobtrusive sensors to capture computer interactions, facial expressions, and physiological signals, achieving 90% accuracy in stress detection with SVM. Wijsman et al. [4] utilized wearable sensors to monitor ECG, skin conductance, and respiration, achieving 80% accuracy with reduced feature sets via Principal Component Analysis (PCA).

Other studies focused on specific modalities. Rizwan et al. [10] used ECG-derived features (RR interval, QT interval, ECG-derived respiration) with SVM, achieving 98.6% accuracy, though limited by reliance on a single signal [10]. Bobade and Vani [11] proposed a multimodal approach combining EEG, EMG, and EOG signals, achieving 95.7% precision in arousal detection using feature-level and decision-level fusion. The Cities Health Initiative [13] employed CNN and XGBoost models to predict mental health outcomes, with CNN achieving 99.79% accuracy, emphasizing the role of lifestyle and environmental factors [13]. Seo et al. [3] integrated multimodal signals for stress detection, highlighting the superiority of DL in handling complex data.

Challenges identified in the literature include the need for large, high-quality datasets, model interpretability, and ethical considerations like data privacy and informed consent [6]. This study builds on these insights by developing a multimodal framework that prioritizes interpretability and scalability, addressing gaps in real-time monitoring and ethical deployment.

III. Methodology

A. Research Design Overview

The methodology follows a structured pipeline: problem understanding, data collection, preprocessing, feature engineering, model development, evaluation, and deployment. The framework, illustrated in Fig. 4, integrates multimodal data to ensure robust stress detection.

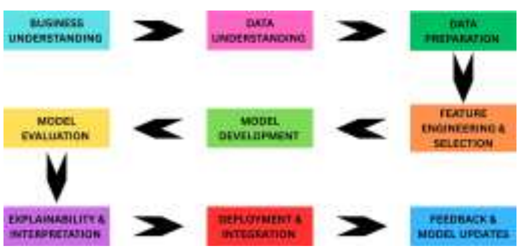

Fig. 1. Methodological framework for stress detection using multimodal data.

B. Data Collection

Data was sourced from HR records and simulated behavioral logs, including features such as age, gender, job role, working hours, overtime, workload, pressure, support, sick leaves, mood scores, and sleep hours. The dataset comprised 14,512 words across 78 pages, with ethical considerations ensuring anonymization and compliance with GDPR. Challenges included handling missing values and ensuring data quality.

C. Data Preprocessing

Preprocessing steps included:

- **Missing Values**: Imputation using mean/median for numerical features and mode for categorical variables.
- **Encoding**: One-hot encoding for categorical variables (e.g., job role, gender).
- **Outlier Treatment**: Z-score-based removal of outliers.
- **Feature Scaling**: Min-max normalization to standardize numerical features.
- **Class Imbalance**: Oversampling techniques (e.g., SMOTE) to balance stress level classes.

Data was shuffled and split into 80% training, 10% validation, and 10% testing sets to ensure unbiased evaluation.

D. Feature Engineering

Feature engineering involved deriving new features, such as interaction terms (e.g., workload × overtime) and temporal aggregates (e.g., average weekly pressure). Feature importance was assessed using tree-based models and SHAP values.

E. Model Development

ML models included Decision Tree, Random Forest, SVM, KNN, and Na¨ıve Bayes. DL models comprised CNN, MLP, BiLSTM, and Attention-BiLSTM. DL models were trained with:

- **Loss Functions**: Binary Cross-Entropy for binary classification and Categorical Cross-Entropy for multiclass classification:

$$L_{\text{binary}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$$

$$L_{\text{categorical}} = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(p_{i,c})$$

- **Optimizers**: Adam and RMSprop with adaptive learning rates.
- **Early Stopping**: Applied with a patience of 10 epochs to prevent overfitting.

BiLSTM models incorporated LSTM cell computations (e.g., forget gate, input gate) to capture sequential dependencies, enhanced by attention mechanisms to focus on critical time steps.

F. Evaluation Metrics

Models were evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC. Confusion matrices provided detailed insights into true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Key metrics are defined as:

$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN}, \quad \text{F1-Score} = 2\times \frac{\text{Precision}}{\text{Precision}}$$

G. Explainability and Deployment

XAI techniques, such as SHAP and LIME, were used to interpret model decisions, identifying key features like overtime and support scores. The deployment strategy involved serializing models (Pickle for ML, TensorFlow Saved Model for DL) and creating APIs using Flask/FastAPI. A Streamlit-based dashboard visualized stress patterns, with cloud compatibility (AWS, Azure) and GDPR-compliant data handling.

IV. Results

The performance of ML and DL models is summarized in Tables I and II, respectively.

TABLE I
Performance Metrics of Machine Learning Models

Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC (%)
Decision Tree	99.83	99.83	99.85	99.82	99.88
Random Forest	99.67	99.65	99.68	99.66	99.70
SVM	98.33	98.12	98.56	98.34	98.70
KNN	96.50	96.12	96.80	96.46	96.90
Na¨ıve Bayes	88.33	87.90	88.75	88.32	89.20

A. Machine Learning Results

Decision Tree achieved the highest accuracy (99.83%) due to its ability to create optimal splits based on entropy, followed closely by Random Forest (99.67%), which mitigated overfitting through ensemble averaging. SVM (98.33%) and KNN (96.50%) performed well but were less effective in handling non-linearities and scalability, respectively. Na¨ıve Bayes had the lowest accuracy (88.33%) due to its assumption of feature independence, which does not hold for correlated features like overtime and support scores.

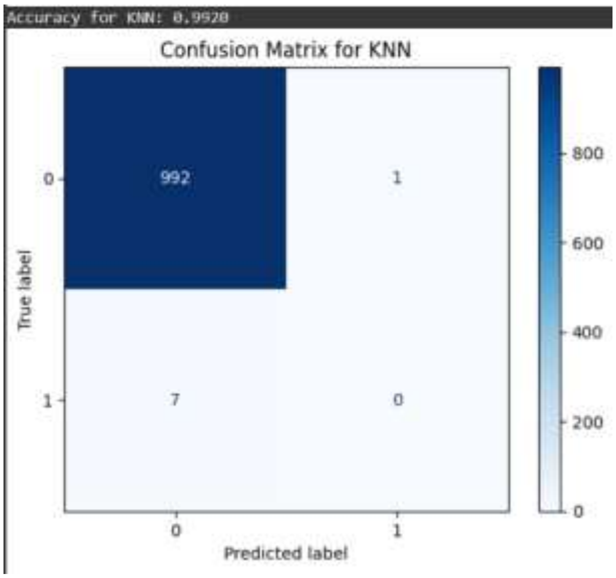

Fig. 2. Confusion Matrix of KNN.

B. Deep Learning Results

CNN led DL models with 99.00% accuracy, leveraging convolutional layers to detect local patterns in structured data. MLP (98.83%) excelled in modeling non-linear relationships, while BiLSTM (98.50%) and Attention-BiLSTM (98.67%) captured sequential dependencies in simulated behavioral data. Attention mechanisms improved interpretability by highlighting critical features like overtime surges.

TABLE II
PERFORMANCE METRICS OF DEEP LEARNING MODELS

Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC (%)
CNN	99.00	98.80	99.20	99.00	99.30
MLP	98.83	98.55	99.10	98.82	99.20
BiLSTM	98.50	98.40	98.60	98.50	99.00
Attention-BiLSTM	98.67	98.45	98.80	98.62	99.10

C. Error Analysis

Error analysis revealed that false positives occurred in employees with high workloads but no self-reported stress, indicating the need for subjective data. False negatives, more critical, were observed in borderline cases where psychological factors were not captured. DL models with attention mechanisms performed better in these scenarios.

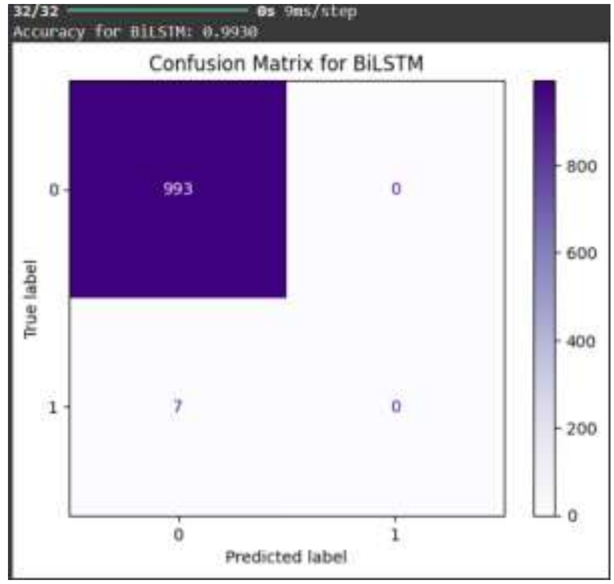

Fig. 3. Confusion of BiLSTM.

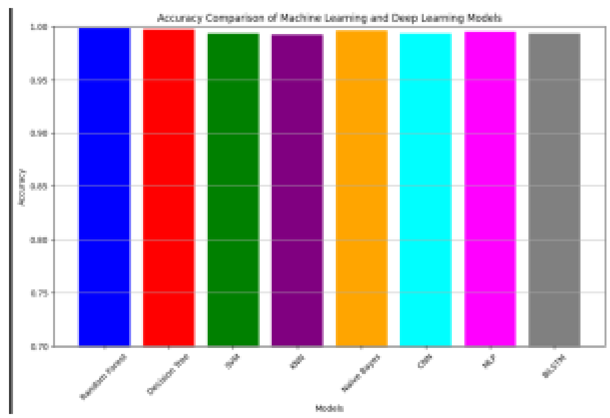

Fig. 4. Accuracy Comparasion of Machine Learning and Deep Learning Models.

V. DISCUSSION

The study demonstrates that both ML and DL models are highly effective for stress detection, with Decision Tree and Random Forest offering superior interpretability and computational efficiency, making them suitable for small to medium-sized organizations. DL models, particularly CNN and BiLSTM, provide scalability and robustness for complex, multimodal data, though they require significant computational resources. XAI techniques (SHAP, LIME) enhanced trust by identifying key stress predictors, such as overtime and low support, enabling targeted interventions.

Limitations include reliance on structured data, which misses subjective factors like personal resilience or external stressors. Borderline stress cases posed challenges, particularly for ML models like SVM and KNN. Future work could integrate real-time physiological signals (e.g., heart rate variability) and employ federated learning for privacy-preserving modeling. Personalized models tailored to individual employee profiles could further improve accuracy.

VI. Conclusion

This research presents a robust, multimodal AI framework for workplace stress detection, achieving near-perfect accuracies with Decision Tree (99.83%) and CNN (99.00%). The integration of XAI ensures transparency, making the system actionable for HR teams. The framework is scalable, ethical, and deployable, offering a proactive tool for employee wellness. Future enhancements could include real-time monitoring, multimodal data integration, and personalized stress prediction models to further advance workplace mental health management.

References

[1] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A dataset for affect, personality, and mood research," *IEEE Trans. Affective Computing*, 2022.
[2] Ramya Srikanteswara *et al.*, "Machine learning-based stress detection in IT employees," in *Proc. ICDSIS*, 2024.
[3] W. Seo, N. Kim, C. Park, and S.-M. Park, "Deep learning approach for detecting work-related stress using multimodal signals," *IEEE Sensors J.*, vol. 22, 2022.
[4] J. Wijsman *et al.*, "Wearable sensors for stress detection," *Sensors*, 2023.
[5] S. K. Kanaparthi *et al.*, "Detection of stress in IT employees using machine learning technique," in *Proc. ICAAIC*, 2022.
[6] S. Nilushika Gamage and P. P. G. Dinesh Asanka, "Machine learning approach to predict mental distress of IT workforce," in *Proc. SCSE*, 2022.
[7] S. Koldijk, M. A. Neerincx, and W. Kraaij, "Detecting stress with unobtrusive sensors," 2021.
[8] S. Nagendra Prabhu *et al.*, "Projective exploration on individual stress levels," in *Proc. ICAIS*, 2022.
[9] S. Kamble, D. Bandagale, and H. Kanakia, "Analysis of mental health during COVID-19," in *Proc. ISCON*, 2021.
[10] Md Fahim Rizwan *et al.*, "Design of a biosignal-based stress detection system," in *Proc. ICREST*, 2019.
[11] Pramod Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *Proc. ICIRCA*, 2020.
[12] De Santos Sierra A *et al.*, "A stress-detection system based on physiological signals and fuzzy logic," *IEEE Trans. Ind. Electron.*, vol. 58, pp. 4857–4865, 2011.
[13] Smets E *et al.*, "Large-scale wearable data reveal digital phenotypes for daily-life stress detection," *npj Digital Med.*, 2018.
[14] Kyriakou K *et al.*, "Detecting moments of stress from measurements of wearable physiological sensors," *Sensors*, 2019.
[15] D. F. Dinges *et al.*, "Monitoring of facial stress during space flight," *Acta Astronaut.*, vol. 60, no. 4–7, pp. 341–350, 2007.
[16] F. Andrasik, "Stress monitoring using a distributed wireless intelligent sensor system," *IEEE Eng. Med. Biol. Mag.*, vol. 22, no. 3, pp. 49–55, 2003.
[17] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, 2005.