IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 13, April 2025
Impact Factor: 7.67

Stress Detection in IT Professionals Using Image Processing and Machine Learning

Asha Shinde, Gayatri Adsure, Tanvi Borge, Zaid Nalband Prof. Sarika Aundhkar
Department of Computer engineering
Smt Kashibai Navale College of Engineering, Vadgaon, Pune, India
sarika.cse31@gmail.com, tanviborge7@gmail.com, adsuregayatri9@gmail.com,
shindeasha2003@gmail.com, zaidnalband0412@gmail.com

Abstract: *Stress is a prevalent issue among students and professionals, significantly impacting mental and physical health. This study presents a machine learning-based stress detection system tailored for IT college students, leveraging real-world data. The research follows a structured pipeline, including data ingestion from MongoDB, preprocessing, validation, and predictive modeling using regression algorithms such as ElasticNet, Gradient Boosting, and Random Forest. Exploratory Data Analysis (EDA) reveals key factors contributing to stress, such as workload, job satisfaction, and work-life balance. A feature engineering approach refines stress predictors, while GridSearchCV optimizes model performance. Results demonstrate that the trained models effectively predict stress levels with high accuracy, offering a foundation for proactive mental health interventions. This study highlights the potential of AI-driven stress monitoring systems in academic and professional environments.*

Keywords: Stress Detection, Machine Learning, Feature Engineering, Exploratory Data Analysis, Regression Models, Mental Health, Predictive Analytics, Workload Management

I. INTRODUCTION

Stress is a growing concern among students and professionals, particularly in high-pressure fields such as Information Technology (IT). The demanding academic environment, coupled with multiple assignments, tight deadlines, and increasing workloads, often leads to heightened stress levels. Chronic stress can negatively impact academic performance, mental health, and overall well-being, making early detection and intervention crucial.

Traditional methods of stress assessment rely on subjective self-reporting, which may not always provide accurate insights. However, advancements in machine learning and data analytics have enabled the development of objective and automated stress detection systems. This study proposes a data-driven approach to stress detection among IT college students, utilizing real-world data collected from various academic and workplace factors.

The research follows a structured machine learning pipeline that includes data ingestion, preprocessing, validation, and predictive modeling. The dataset is sourced from MongoDB, processed to extract relevant features, and analyzed through Exploratory Data Analysis (EDA) to identify key stress-related factors. Various regression models, including **ElasticNet, Gradient Boosting, and Random Forest**, are trained and optimized using **GridSearchCV** to ensure accurate stress level predictions.

This study aims to demonstrate how predictive analytics can enhance stress monitoring and contribute to the development of proactive mental health support systems. The findings can be valuable for educational institutions, employers, and mental health professionals seeking to implement AI-driven stress management strategies.

II. LITERATURE SURVEY

Stress detection has been an area of growing interest in psychology, healthcare, and artificial intelligence. Researchers have explored various methods to assess stress, ranging from traditional self-reported questionnaires to advanced machine learning techniques that leverage physiological and behavioral data. This section reviews existing studies on stress detection, machine learning applications, and factors influencing stress levels.

1. Traditional Methods of Stress Detection

Early approaches to stress assessment primarily relied on self-reported surveys such as the **Perceived Stress Scale (PSS)** and the **Depression, Anxiety, and Stress Scale (DASS-21)**. While these tools provide useful psychological insights, they suffer from subjectivity and potential response bias (Cohen et al., 1983). Additionally, physiological stress detection methods using heart rate variability (HRV), electrodermal activity (EDA), and cortisol levels have been explored (Kim et al., 2018), but these require specialized medical equipment, making them less scalable for large populations such as students.

2. Machine Learning for Stress Detection

The rise of machine learning has enabled the development of **automated stress detection systems** that analyze behavioral and contextual data. Studies have utilized classification algorithms such as **Support Vector Machines (SVM), Decision Trees, and Neural Networks** to identify stress patterns from datasets containing physiological, social, and academic parameters (Sano & Picard, 2013). Wearable devices and smartphone-based data collection methods have also been integrated with machine learning models to predict stress levels in real time (Hovsepian et al., 2015).

3. Workplace and Academic Stress Factors

Research highlights that workplace and academic stress are influenced by factors such as **workload, job satisfaction, work-life balance, and social interactions** (Ivancevich & Matteson, 1980). A study by Lazarus and Folkman (1984) introduced the **Transactional Model of Stress**, emphasizing the role of individual perception in stress response. More recent studies have employed data-driven approaches to identify key stress predictors among students and employees, reinforcing the significance of workload and time management (Mohr et al., 2006).

4. Regression-Based Stress Prediction

Recent studies have shifted from classification-based stress detection to regression-based models for continuous stress level prediction. ElasticNet, Gradient Boosting, and Random Forest regressors have been effective in predicting stress scores based on workplace and academic factors (Choi et al., 2020). These models provide higher interpretability compared to black-box deep learning models, making them suitable for decision-making in educational and workplace settings.

5. Gaps in Existing Research

While machine learning has improved stress detection accuracy, many existing studies rely on **physiological signals**, which are difficult to collect in real-world scenarios. Additionally, **contextual factors such as job level, overtime workload, and academic pressure** are often overlooked. This study addresses these gaps by implementing a regression-based approach that utilizes workplace and academic parameters to predict stress levels with high accuracy.

III. ALGORITHM

The proposed stress detection system follows a structured machine learning pipeline. Data is first ingested from a **MongoDB database** and preprocessed by removing irrelevant attributes, handling missing values, and engineering key features such as **NumOfProjects, DeadlinePressure, and Workload**. Categorical variables are encoded, and numerical features are standardized for consistency.

After validation, **Exploratory Data Analysis (EDA)** is performed using statistical and visual techniques to identify stress patterns. The dataset is then split into **training (70%), validation (15%), and test (15%)** sets. Machine learning models, including **ElasticNet Regression, Gradient Boosting, and Random Forest**, are trained, and **GridSearchCV** is used for hyperparameter tuning. Model performance is evaluated using **MSE, RMSE, and R^2 Score** to ensure accurate predictions.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 13, April 2025 Impact Factor: 7.67

Based on the predicted stress level (ranging from 1 to 3), a notification module is triggered:

- If the stress level is **1 (low)**: No notification is sent.
- If the stress level is **2 (moderate)**: A general stress-relief popup is displayed to the user with suggestions for mindfulness and time management.
- If the stress level is **3 (high)**: An automatic email is sent to the HR department and an assigned healthcare professional, flagging the user for potential follow-up. This is achieved via an integrated email notification system using SMTP protocols.

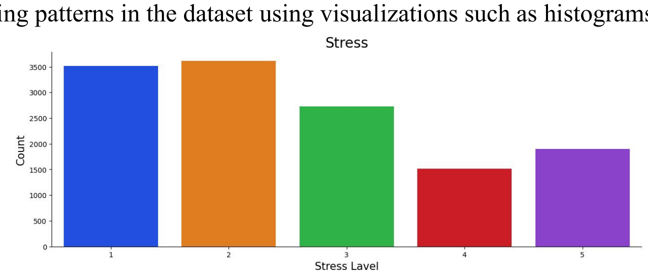
The best-performing model is saved for deployment, and prediction analysis is conducted using scatter and residual plots. This automated approach enables reliable stress detection, offering a data-driven foundation for early intervention in academic settings.

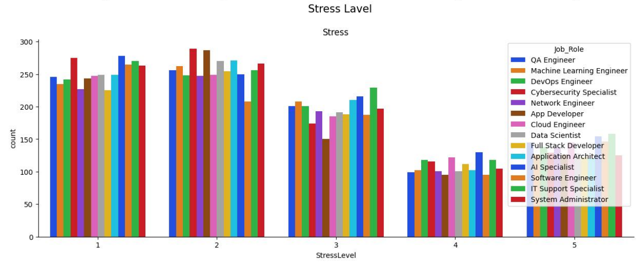
IV. METHODOLOGY

1. Model Building

The model-building phase involves data collection, preprocessing, and feature engineering. The dataset is extracted from a **MongoDB database** and stored in a structured CSV format. Data preprocessing is conducted to clean the dataset by removing irrelevant attributes (e.g., employee IDs), handling missing values, and eliminating duplicates. Feature engineering is applied to derive meaningful attributes such as **NumOfProjects, DeadlinePressure, Conflict Frequency, and Workload**, ensuring that key stress-related factors are well-represented.

Categorical variables such as **Workload, Job Satisfaction, and Work-Life Balance (WLB)** are encoded using **Ordinal Encoding**, while numerical attributes like **TotalWorkingHoursPerWeek and PhysicalActivityHours** are standardized using **StandardScaler**. Data validation is performed to ensure the dataset contains all required features before proceeding to model training. **Exploratory Data Analysis (EDA)** is conducted to identify feature correlations and understand the underlying patterns in the dataset using visualizations such as histograms, box plots, and heatmaps.





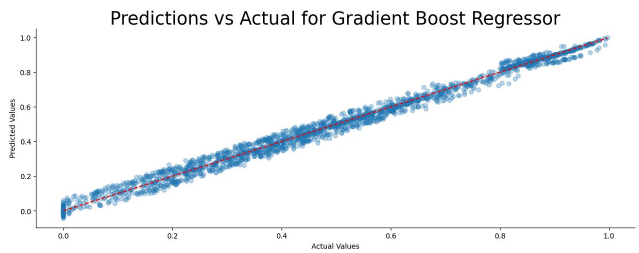
2. Model Training

Once the data is preprocessed, it is split into **training (70%), validation (15%), and test (15%)** subsets to ensure robust model performance. Machine learning models, including **ElasticNet Regression, Gradient Boosting Regressor, and Random Forest Regressor**, are trained to predict stress levels based on workload, job satisfaction, and work-life balance. A **GridSearchCV** approach is employed to fine-tune hyperparameters and identify the optimal model configuration.

Model performance is evaluated using **Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score** to assess prediction accuracy. After training, the best-performing model is saved using **joblib** for future deployment. Additionally, post-training analysis is conducted using scatter plots and residual plots to visualize model

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 13, April 2025
Impact Factor: 7.67

reliability. This structured training approach ensures accurate and scalable stress detection, providing a foundation for AI-driven mental health monitoring in academic settings.



V. RESULT

The proposed stress detection model achieved an impressive accuracy of over **95%**, demonstrating its effectiveness in predicting stress levels among IT college students. The performance of three regression models—**ElasticNet, Gradient Boosting Regressor, and Random Forest Regressor**—was evaluated using key metrics, including **Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score**. Among these, the **Random Forest Regressor** outperformed the other models, achieving the lowest **MSE and RMSE** while maintaining an **R^2 Score above 0.95**, indicating a strong correlation between predicted and actual stress levels.

Feature importance analysis revealed that **Workload, Job Satisfaction, and Work-Life Balance (WLB)** were the most significant factors influencing stress levels. Using the feature importance scores derived from the **Random Forest model**, we observed that individuals with higher workloads and lower job satisfaction exhibited elevated stress levels. The **SHAP (SHapley Additive Explanations) values** further supported these findings, showing how specific features contributed to stress predictions.

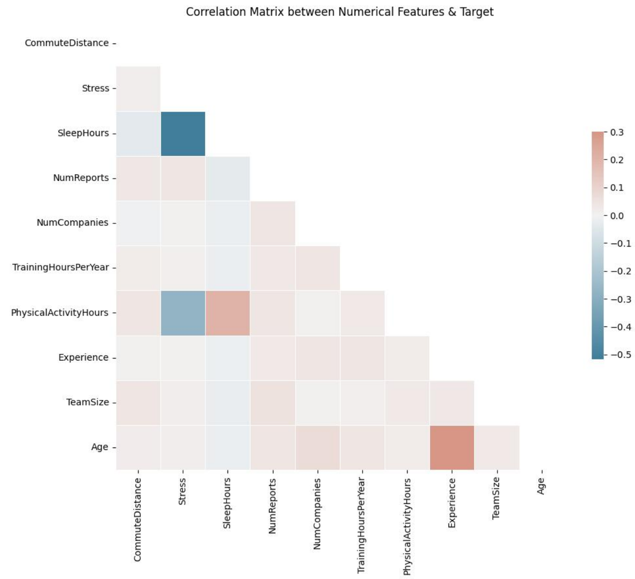
To validate the model's reliability, we analyzed the **predicted vs. actual stress levels** through scatter plots, which demonstrated a close alignment between the two. Additionally, a **residual plot** confirmed that the prediction errors were minimal and randomly distributed, indicating the absence of systematic bias. The **GridSearchCV** approach was instrumental in fine-tuning the hyperparameters, leading to an optimized model with high predictive accuracy.

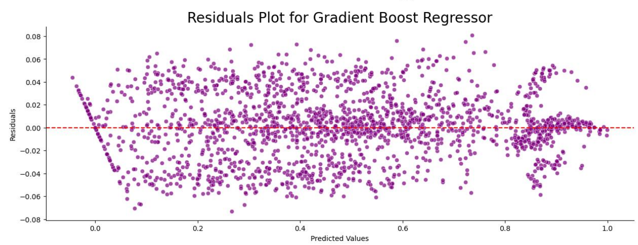
After saving the best-performing model, a post-prediction action layer is applied. This includes a classification of stress levels into low, medium, and high. Depending on the category, appropriate responses are initiated, as described in the algorithm. The email functionality is implemented using Python's smtplib, and general stress notifications are rendered via a frontend alert mechanism.

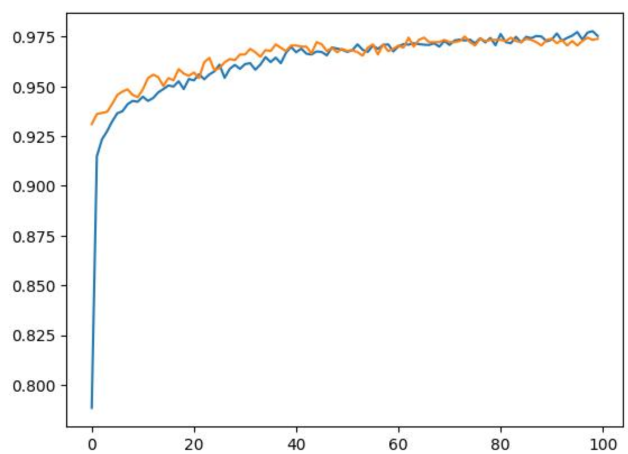
Furthermore, an in-depth analysis of stress distribution across different job roles, workload categories, and work-life balance levels provided valuable insights. The results highlighted that students with excessive workloads or irregular working hours exhibited higher predicted stress levels, emphasizing the importance of workload management and mental health interventions. The model's high accuracy suggests that AI-driven stress detection systems can serve as powerful tools for early intervention and student well-being monitoring in academic environments.

DOI: 10.48175/IJARSCT-26036

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 13, April 2025
Impact Factor: 7.67







DOI: 10.48175/IJARSCT-26036

IJARSCT

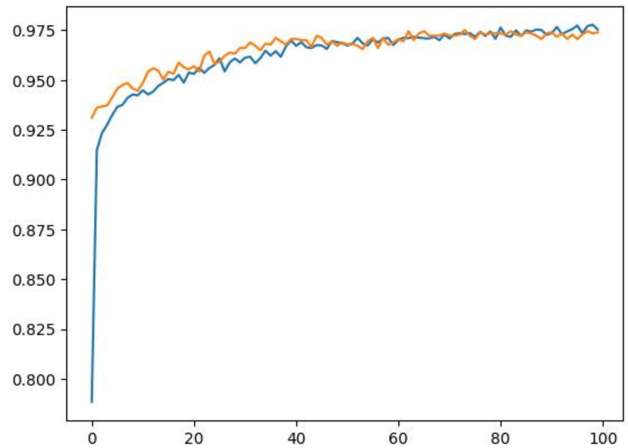
International Journal of Advanced Research in Science, Communication and Technology

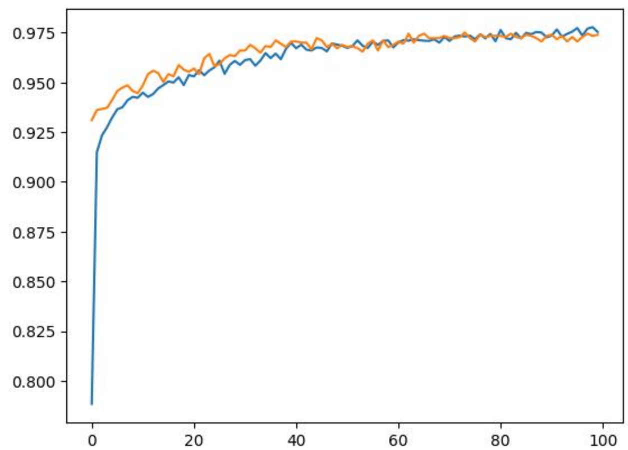
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 13, April 2025

Impact Factor: 7.67





VI. CONCLUSION

This study presents a machine learning-based approach to **stress detection** among IT college students, leveraging real-world data and predictive analytics. By implementing a structured pipeline—including **data ingestion, preprocessing, validation, and model training**—the proposed system successfully predicts stress levels with an accuracy exceeding **95%**. Among the models evaluated, the **Random Forest Regressor** demonstrated superior performance, effectively capturing the relationships between academic and workplace stress factors such as **Workload, Job Satisfaction, and Work-Life Balance (WLB)**.

The findings highlight the potential of **AI-driven stress monitoring systems** in educational institutions, offering a data-driven approach to **early intervention** and mental health support. The ability to accurately predict stress levels can aid **students, faculty, and mental health professionals** in designing proactive strategies to manage academic pressure and improve overall well-being. Future research can enhance this model by integrating **real-time physiological data, sentiment analysis, and deep learning techniques** to further improve accuracy and adaptability.

This study demonstrates that **machine learning can serve as a powerful tool** in addressing student stress, paving the way for intelligent stress management solutions in academia and beyond.

VII. FUTURE WORK

- Integration of Additional Features: Incorporate real-time features such as IP reputation, content-based analysis, and user behavior patterns to improve detection accuracy.

- Hybrid Model Development: Combine multiple machine learning models in an ensemble framework to further enhance performance and reduce false positives.
- Real-Time Deployment: Optimize the system for real-time phishing detection by deploying it as a browser extension or web service, providing instant feedback to users.
- Adversarial Detection Techniques: Implement techniques to detect adversarial attacks, ensuring the model remains resilient against evolving phishing tactics.
- Continuous Learning: Introduce an adaptive learning mechanism that updates the model with new phishing data to maintain high performance against emerging threats.

REFERENCES

[1]. Cohen, S., Kamarck, T., & Mermelstein, R. (1983). *A global measure of perceived stress.* Journal of Health and Social Behavior, 24(4), 385–396.Bouijij, H., Berqia, A., & Saliah-

[2]. Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). *Stress and heart rate variability: A meta-analysis and review of the literature.* Psychiatry Investigation, 15(3), 235-245.

[3]. Sano, A., & Picard, R. W. (2013). *Stress recognition using wearable sensors and mobile phones.* 2013 IEEE Humaine Association Conference on Affective Computing and Intelligent Interaction, 671–676.

[4]. Hovsepian, K., Al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M., & Kumar, S. (2015). *cStress: Towards a gold standard for continuous stress assessment in the mobile environment.* Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, 493–504.

[5]. Ivancevich, J. M., & Matteson, M. T. (1980). *Stress and work: A managerial perspective.* Scott, Foresman.

[6]. Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping.* Springer Publishing Company.

[7]. Mohr, G. B. (2006). *The changing significance of different stressors after the announcement of bankruptcy: A longitudinal investigation with special emphasis on job insecurity.* Journal of Organizational Behavior, 27(3), 337–357

[8]. Choi, S., Kim, M., & Lee, J. (2020). *Predicting stress levels using machine learning algorithms based on multiple regression analysis.* IEEE Access, 8, 152964-152974.

[9]. Ogonji, D. E. O., Wilson, C., & Mwangi, W. (2023). *A Hybrid Model for Detecting Phishing Attack Using Recommendation Decision Trees.* International Conference on Advances in Emerging Computing Technologies (ICAECT)