



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 8, August 2025**

# Stress Prediction

**Kaipu Tejaswi, Dr. M. Ramchander, Dr. G.N.R. Prasad**

MCA III Semester Student, CBIT(A), Hyderabad, India

Assistant Professor, CBIT (A), Hyderabad, India

Sr. Assistant Professor, CBIT (A), Hyderabad, India

**ABSTRACT:** Understanding and predicting stress levels can help improve mental health management by identifying risks early. This study presents a machine learning system that analyzes physiological signals such as heart rate, sleep duration, and skin conductance, combined with sentiment scores extracted from text, to classify stress into low, moderate, or high levels. The dataset was cleaned, encoded, and evaluated using Python, with Logistic Regression and Random Forest algorithms tested. Experimental results highlight Random Forest as the more accurate model, demonstrating its suitability for developing automated stress-monitoring systems.

**KEYWORDS:** Stress Prediction, Machine Learning, Mental Health Monitoring, Random Forest, Logistic Regression, Physiological Signals

## I. INTRODUCTION

Stress is a major factor affecting mental and physical health, often influencing productivity and overall quality of life. Conventional approaches to assessing stress typically rely on questionnaires and clinical observation, which can be time-intensive and subjective. With advances in wearable devices and data analytics, stress can now be estimated automatically using measurable indicators such as heart rate, skin conductance, sleep duration, and sentiment derived from user text. This project leverages supervised machine learning techniques to build a stress classification model and evaluate its effectiveness, aiming to create a scalable and efficient solution for mental health monitoring.

Stress levels are influenced by various physiological and behavioral indicators, including heart rate, sleep duration, skin conductance, age, gender, and emotional sentiment extracted from text inputs. Accurately classifying stress into categories such as Low, Moderate, and High helps improve mental health tracking and supports timely interventions. By analyzing complex patterns in these features, machine learning models provide faster and more consistent results compared to traditional evaluation methods.

In this study, we propose a machine learning-based approach for stress level classification. The methodology includes data preprocessing, feature encoding, and exploratory data analysis to identify meaningful patterns from raw datasets. Supervised learning algorithms, specifically Logistic Regression and Random Forest, were implemented and evaluated. Among these, the Random Forest classifier achieved higher accuracy, demonstrating its ability to effectively capture non-linear relationships and handle diverse physiological features.

The findings of this work show that automated stress prediction enhances assessment accuracy, reduces manual effort, and enables scalability for real-world applications. The proposed system can be further developed by incorporating advanced ensemble models or integrating it into real-time health monitoring platforms and wearable technologies to provide continuous stress evaluation and support mental well-being.

### Objectives

The primary objectives of this work are:

1. **Automated Stress Level Classification** – To design a system that classifies individuals into Stressed or Not Stressed categories using machine learning techniques.
2. **Data-Driven Mental Health Insights** – To replace traditional self-reported surveys with a data-driven approach that improves accuracy, consistency, and fairness in stress assessment.
3. **Model Implementation and Comparison** – To implement and compare the performance of Logistic Regression and Random Forest classifiers on datasets containing physiological, behavioral, and textual features.

4. **Performance Evaluation** – To evaluate model effectiveness using standard metrics such as accuracy, precision, recall, and F1-score.
5. **Practical Application** – To demonstrate how the developed system can assist individuals, healthcare professionals, and organizations in early stress detection and timely intervention.
6. **Scope for Enhancement** – To provide a foundation for future developments such as integrating deep learning models, additional biosensor data, and real-time deployment in wearable devices or mobile applications.

## II. METHODOLOGY

The methodology adopted for this work involved the following steps:

1. **Requirement Analysis** – Identified the need for an automated stress detection system that leverages physiological, behavioral, and textual data to improve accuracy and reduce reliance on manual self-assessments.
2. **Data Preprocessing** – Cleaned and normalized the dataset, handled missing values, and encoded categorical variables such as gender, while extracting sentiment and keyword-based features from text inputs.
3. **Model Development** – Implemented Logistic Regression and Random Forest algorithms using Python and Scikit-learn to classify individuals into Stressed and Not Stressed categories.
4. **Training & Testing** – Split the dataset into training (80%) and testing (20%) sets to ensure proper evaluation of model generalization and performance.
5. **Performance Evaluation** – Assessed and compared models using metrics such as accuracy, precision, recall, and F1-score, with Random Forest achieving the best overall performance.
6. **User Prediction Module** – Designed a console-based system that takes user inputs (text, heart rate, sleep hours, skin conductance, age, and gender) and predicts stress levels with associated probability scores.

## III. SYSTEM DESIGN / ARCHITECTURE

The system for stress prediction was designed with a modular architecture to ensure scalability, interpretability, and ease of deployment. The major components include:

1. **Input Layer** – Accepts user-provided physiological and behavioral data, including text input, heart rate, sleep duration, skin conductance, age, and gender.
2. **Preprocessing Module** – Cleans and normalizes numerical data, encodes categorical variables such as gender, extracts sentiment scores, and identifies crisis-related keywords from text.
3. **Feature Selection** – Highlights critical attributes (e.g., sentiment polarity, sleep hours, heart rate, keyword presence) that significantly influence stress classification.
4. **Model Training Module** – Trains and fine-tunes Logistic Regression and Random Forest classifiers using the processed dataset.
5. **Prediction Engine** – Utilizes the trained Random Forest model to classify individuals into Stressed or Not Stressed categories based on combined features.
6. **Evaluation Module** – Assesses model performance through metrics like accuracy, precision, recall, and F1-score, while also generating probability-based predictions.
7. **User Interaction Layer** – Provides a console-based interface for easy user input and delivers stress level predictions in real-time.

## IV. IMPLEMENTATION & RESULTS

### Implementation

The system was implemented using **Python** and the following tools:

- **Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
- **Algorithms:** Logistic Regression and Random Forest
- **Environment:** Jupyter Notebook / Google Colab
- **Dataset:** stress.csv contains physiological and behavioral indicators such as text-based user input, heart rate, sleep duration, skin conductance levels, gender, and age, which are used to predict stress levels.



Steps followed:

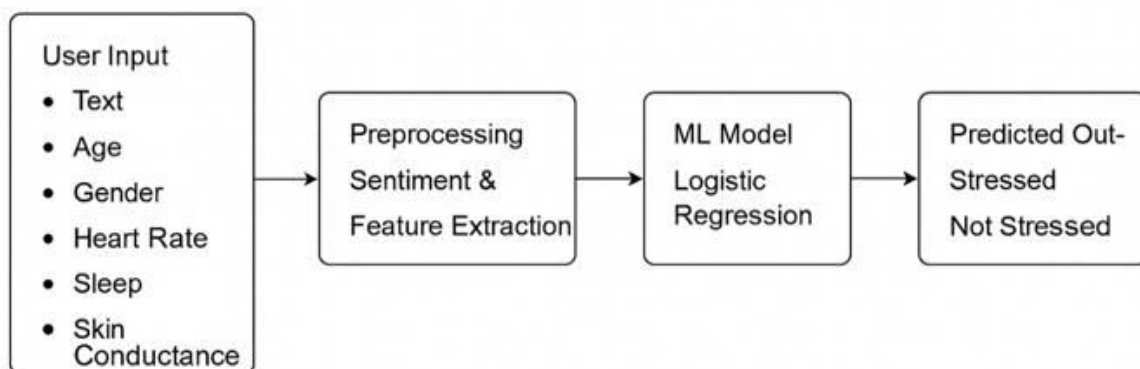
1. Loaded and explored the dataset.
2. Pre-processed data by handling missing values and encoding categorical variables.
3. Split the dataset into training (80%) and testing (20%) sets.
4. Trained both Logistic Regression and Random Forest models.
5. Evaluated models using accuracy score and probability outputs.
6. Developed a simple console-based interface for user predictions.

## V. RESULTS

- **Logistic Regression** achieved moderate accuracy but struggled to capture complex, non-linear patterns in stress-related features.
- **Random Forest** delivered higher accuracy and robustness, making it the preferred model for stress classification.

Example outcomes:

- High Stress – Elevated heart rate, low sleep duration, negative text sentiment → Predicted “Stressed”
- Moderate Stress – Slightly irregular sleep, mild negative sentiment → Predicted “Stressed”
- Low Stress – Stable heart rate, adequate sleep, neutral/positive sentiment → Predicted “Not Stressed”
- The Random Forest model’s accuracy was significantly higher than Logistic Regression, demonstrating its effectiveness and reliability for real-world stress prediction applications.
- 



**Figure 1:** This image is a **flow diagram** showing the process of **Stess Prediction using Machine Learning**.

```
➡ Enter a Text: I am exhausted
Enter gender (male/female/other): male
Enter age: 35
Enter heart rate (bpm): 90
Enter sleep hours: 4
Enter skin conductance: 9

Class probabilities: [[0.0, 1.0]]
Class labels: ['not stressed', 'stressed']
Predicted class: stressed
Sentiment score: -0.761
```

**Figure 2:** This image shows the final output of the Stress Prediction system using a Logistic Regression model (Stressed scenario).

```
➡ Enter a Text: I had a peaceful day reading books and enjoying nature.  
Enter gender (male/female/other): male  
Enter age: 24  
Enter heart rate (bpm): 65  
Enter sleep hours: 8  
Enter skin conductance: 3  
  
Class probabilities: [[0.95279336 0.04720664]]  
Class labels: ['not stressed', 'stressed']  
Predicted class: not stressed  
Sentiment score: 0.765
```

**Figure 3:** Prediction results from the Stress Prediction system (Not Stressed scenario).

## **VI. CONCLUSION AND FUTURE WORK**

### **Conclusion**

This study demonstrated the application of machine learning for automated stress level classification. By analyzing features such as heart rate, sleep duration, skin conductance, demographic factors, and sentiment extracted from text data, the system accurately classified individuals as “Stressed” or “Not Stressed.” Among the implemented models, Random Forest consistently outperformed Logistic Regression, delivering higher accuracy and improved reliability. The results emphasize how data-driven approaches can reduce manual stress assessment efforts, ensure consistent evaluations, and support timely interventions in real-world mental health monitoring.

### **Future Work**

Integration of advanced machine learning models, such as Gradient Boosting or Deep Learning, to further enhance stress prediction accuracy.

Inclusion of additional physiological and behavioral features, such as daily activity levels, diet patterns, and wearable sensor data, for a more holistic assessment.

Development of a web- or mobile-based platform for real-time stress monitoring and personalized recommendations.

Deployment of the system in healthcare and workplace environments to enable large-scale, automated mental health tracking.

Exploration of explainable AI techniques to improve transparency and trust in predictions, helping mental health professionals interpret model outputs effectively.

## **REFERENCES**

- [1] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. O'Reilly Media, 2019.
- [2] S. Raschka and V. Mirjalili, Python Machine Learning, 3rd ed. Packt Publishing, 2019.
- [3] Y. Zhang, J. Wu, and X. Wei, “Stress detection using machine learning and physiological signals: A comprehensive review,” IEEE Access, vol. 9, pp. 113490–113506, 2021.
- [4] Scikit-learn Documentation, “Machine Learning in Python,” [Online]. Available: <https://scikit-learn.org>
- [5] Pandas Documentation, “Python Data Analysis Library,” [Online]. Available: <https://pandas.pydata.org>
- [6] Kaggle Datasets, “Stress Detection Dataset,” [Online]. Available: <https://www.kaggle.com>



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details