Journal of Applied Science, Engineering, Technology and Management
Vol. 4, Issue 1, June 2026 pp. 31 - 34

Predictive Modeling of Stress Levels Using Physiological and Lifestyle Factors

Parvathy R
Dept. of Electrical and Electronics Engineering
Federal Institute of Science And Technology
Kerala, India, Pin: 683577
parvathyrani@fisat.ac.in
ORCID ID: 0000-0001-5822-5526

Unni Kartha G
Dept. of Civil Engineering
Federal Institute of Science And Technology
Kerala, India, Pin: 683577
unnikartha@fisat.ac.in
ORCID ID: 0000-0002-7809-1227

***Abstract*—Stress has emerged as a critical factor that affects both physical and mental health in today's fast paced lifestyle. Early detection of stress levels can enable timely interventions and preventive healthcare. This study presents a data driven approach for stress level classification using physiological and lifestyle features derived from a publicly available dataset. Key variables include age, occupation, cholesterol level, sleep quality, physical activity, and other health related metrics. A Random Forest Classifier is used to model and classify individuals into three stress levels, i.e., low, moderate, and high. The model achieves a classification accuracy of 0.74, with an averaged F1 score of 0.71, indicating a strong generalization performance. Feature importance analysis reveals that occupation, age, and cholesterol level are the most influential predictors of stress. Deeper analysis shows that individuals in high response professions, within the age range of 30 to 40 years, and with elevated cholesterol levels are more prone to high stress. Moreover, the model offers insight into the relative importance of each feature, enhancing the interpretability and potential clinical utility. The proposed model demonstrates that integrating machine learning with wearable and lifestyle data serves as a powerful tool for proactive stress management and personalized healthcare.**

Keywords* — *Stress Prediction, Random Forest Classifier, Physiological Factors, Lifestyle Factors, Machine Learning in Healthcare, Feature Importance Analysis

I. Introduction

In recent years, the integration of Internet of Things (IoT) technologies with wearable devices has revolutionized healthcare by enabling real time monitoring and predictive health analytics. Traditional wearable devices, such as smart watches and fitness trackers, offer limited functionality, high costs, and often lack adaptability to predictive insights, particularly in stress management. Chronic stress, which can lead to cardiovascular diseases, anxiety disorders, and weakened immunity, remains under monitored due to the lack of continuous, affordable, and intelligent monitoring systems. A key innovation of the HOT Watch lies in its machine learning capabilities, particularly in detecting stress based on user health data. Using a publicly available stress detection dataset, a machine learning analysis is performed to identify the most critical features that influence stress. The Random Forest model revealed that occupation, age, and cholesterol Level are dominant indicators of stress. This paper presents the machine learning methodology, and detailed insights into stress prediction, thereby demonstrating the feasibility of HOT Watch for intelligent health monitoring.

II. Literature Review

Wearable health monitoring systems have gained considerable attention in recent years due to their potential for real-time tracking of physiological parameters and early detection of health conditions [1]. Traditional wearables like Fitbit, Apple Watch, and Samsung Galaxy Watch monitor basic vitals such as heart rate and activity levels, but most lack advanced analytics capabilities and fail to provide contextual health insights such as stress prediction.

Recent research has explored the use of machine learning (ML) techniques in wearable-based health monitoring [2], [3]. Use of physiological signals such as heart rate variability and skin temperature to detect stress levels using Support Vector Machines and Decision Trees are mentioned in literature [4]–[7]. Their results showed moderate success, but the systems were not optimized for deployment in resource-constrained wearable devices. Similarly, studies were performed to see the effect of contextual data (location, movement) with physiological metrics to improve stress prediction accuracy, but it shows that these require complex and non scalable hardware [8].

In the domain of IoT-based health monitoring, wearable sensor network architecture was suggested to monitor elderly patients remotely [9]. However, the work focused more on data transmission and less on intelligent decision-making. Attempts were made to incorporate AI into wearable systems for cardiovascular risk detection, demonstrating how real-time analytics could improve preventive care [10]. However, they lacked focus on stress, which remains a critical mental and physical health factor.

A few studies have systematically identified and validated dominant features affecting stress [11], [12]. While most wearable systems simply log raw data, they do not analyze patterns or provide actionable insights. Moreover, there is limited exploration into affordable, customizable, and energy-efficient wearables that can be adapted for various age groups and occupational risks.

© eISSN 2584-0371
2026 Federal Institute of Science and Technology
(FISAT), Angamaly, Kerala, India.

Received for review: 22-05-2026
Accepted for publication: 11-06-2026
Published: 15-06-2026

This paper addresses how machine learning models can predict stress levels induced due to various factors. It further identifies Occupation, Age, and Cholesterol Level as the most significant features contributing to stress, making it more actionable for personalized healthcare. This paper also proposes the development of a low cost, machine learning enabled, and user friendly system capable of collecting physiological data from individuals and utilizing a pre trained model to predict their stress levels.

III. System Design of HOT Watch

The HOT Watch (Health-Oriented Tracker Watch) is a compact, wearable health monitoring device designed for continuous stress detection using physiological signals [13]. The system integrates low power sensors, IoT connectivity, and intelligent algorithms to deliver real time, context aware health monitoring [14]. The architecture of HOT Watch consists of four key layers:

1) *Sensor Layer:* Gathers raw physiological data using onboard biomedical sensors.
2) *Processing Layer:* Performs signal preprocessing, feature extraction, and classification using embedded or cloud-based ML models.
3) *Communication Layer:* Uses IoT protocols (e.g., MQTT, HTTP) to transmit data securely to cloud or mobile applications.
4) *Application Layer:* Visualizes the user's stress level and alerts healthcare providers or caregivers when abnormal readings are detected.

IV. Machine Learning Model - Random Forest Classifier

Random Forest is a powerful supervised machine learning algorithm widely used for classification and regression tasks. It belongs to the family of ensemble methods, which aim to combine multiple individual models (called base learners) to produce a more accurate and robust overall model [15]–[17]. The fundamental idea behind Random Forest is to build a "forest" of decision trees during training and output the class that is the mode (majority vote) of the individual trees' predictions. It mitigates the tendency of decision trees to overfit on the training data by introducing two layers of randomness. Each tree is trained on a randomly selected subset of the training data, with replacement. At each split in the tree, a random subset of features is considered, rather than evaluating all possible features. This randomization ensures that individual trees are decorrelated, and their errors are less likely to overlap, improving overall performance [18].

One of the key strengths of this approach lies in its non parametric nature and its ability to estimate feature importance [19]–[21]. Being non-parametric means that the classifier does not make any assumptions about the underlying distribution of the data, making it highly flexible and suitable for a wide range of real-world problems [22]. Additionally, Random Forest provides a mechanism to evaluate the relative importance of input features by analyzing how much each variable contributes to the predictive performance of the model [23], [24].

Although Random Forest is a well established algorithm, the novelty of this work lies in its application to a three-class stress detection problem (low, medium, high) using physiological and behavioral features from the Kaggle dataset. Random Forest was chosen for its robustness, interpretability, and capability to rank feature importance, which is crucial for understanding stress related predictors. This work thus emphasizes both the practical applicability and interpretability of the approach rather than algorithmic novelty, providing a foundation for future research on more advanced models and enriched datasets.

V. Machine Learning Model Design

The core intelligence of the HOT Watch system lies in its ability to predict stress levels using machine learning (ML) algorithms based on physiological and demographic data. This section presents the feature selection and model evaluation in model development process.

A. Data Description

The machine learning model was trained using a publicly available Stress Detection in Students Dataset from Kaggle, consisting of 733 entries whose attributes are as categorised as follows.

1) *Demographic:* Age, Occupation, Gender, Marital Status.
2) *Health condition:* Cholesterol Level, Blood Sugar Level, Blood Pressure.
3) *Lifestyle:* Physical Activity, Sleep Duration, Sleep Quality, Caffeine Intake, Screen Time, Exercise Type, Meditation Practice, Social Interactions, Alcohol Intake, Smoking Habit, Work Hours, Travel Time.
4) *Target Variable:* Stress Detection (Low, Medium, High).

B. Data Preprocessing

The following steps were included before using the data for learning purpose.

1) *Handling Missing Values:* Rows with missing target labels were removed.
2) *Label Encoding:* This includes target label encoding and categorical feature encoding. The target stress levels are encoded as 'High': 0, 'Low': 1, 'Medium': 2. In categorical encoding, variables like Gender, Occupation, Marital Status, Smoking Habit, Meditation Practice, Exercise Type were encoded numerically.
3) *Drop irrelevant columns:* Time-based columns like Wake Up Time, Bed Time are hard to interpret without extra parsing. Hence they are omitted.
4) *Training - testing data split:* The data is splitted into 80% for training and 20% for testing.
5) *Feature Scaling:* Normalization of continuous features using Min-Max scaling.

C. Feature Selection and Feature Importance Identification

A Random Forest Classifier was used to compute the feature importance scores. The top contributors to stress level prediction are identified as age, occupation and cholesterol level in Fig. 1.

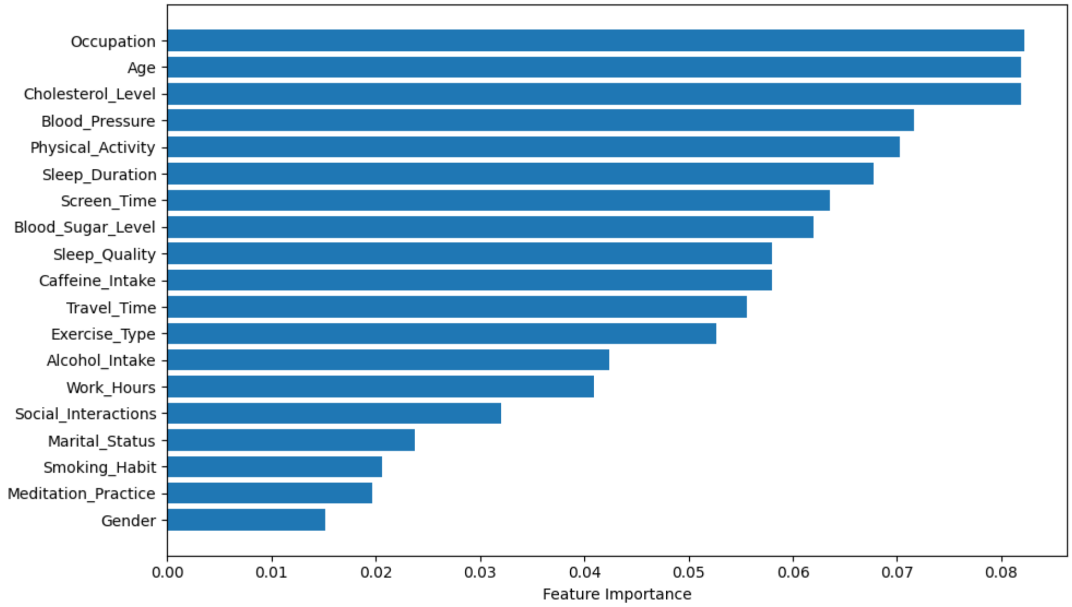

Fig. 1. Feature Importance from Random Forest Classifier

TABLE I: PERFORMANCE METRICS FOR STRESS LEVEL CLASSIFICATION

Stress Level	Precision	Recall	F1-score	Support
Low (0)	0.85	0.78	0.82	60
Medium (1)	0.71	0.45	0.56	33
High (2)	0.67	0.85	0.75	62

VI. RESULTS

This section presents the experimental results of applying the Random Forest classifier to the stress detection dataset. The model was trained and evaluated to classify user stress levels into three categories: low (0), medium (1), and high (2), based on physiological and lifestyle features.

A. Model Performance Overview

The Random Forest classifier achieved an accuracy of 0.74 and F1 Score of 0.71.

The classification model performed well in identifying high and low stress levels, but the moderate class showed weaker results (Table I). This can be attributed to the overlap of the feature distributions between the moderate and the other two categories. This overlap makes the model less confident in distinguishing moderate stress instances. Since the data set was sourced from Kaggle, no additional data is currently available to expand or balance the classes. This limitation restricts further improvement in the present study. Future work could explore advanced techniques such as hierarchical classification, ensemble approaches, or synthetic data generation to enhance moderate class recognition.

B. Top Contributing Features

Using the Random Forest's feature importance method, the following were identified as dominant stress predictors: (a) occupation, (b) age and (c) cholesterol level.

C. Detailed Category-Wise Risk Analysis

The average feature value for various stress levels are tabulated below in Tables II–VI. Following inferences are made after evaluating the experimental outcomes for the features selected.

1) *Occupation:* High-Stress Group (Stress Level 2) had an average occupation code of 90.3. Moderate-Stress Group had a higher average of 96.4, indicating that certain demanding or high-responsibility occupations may lead to stress. Low-Stress Group had the lowest occupation code average (78.3), possibly linked to more flexible or lower-pressure professions.
2) *Age:* The Low-Stress Group had the highest average age: 41.5 years. Moderate and High-Stress groups had younger averages: 36.1 and 37.8 years, respectively. Older individuals (possibly due to experience or lifestyle stability) tend to report lower stress, while younger adults in their mid 30s are more prone to stress likely due to career and life transitions.
3) *Cholesterol Level:* Low stress individuals had higher average cholesterol, suggesting that cholesterol may not directly correlate with stress in a linear way. However, extreme fluctuations, especially in the moderate stress group, could indicate poor health management practices under stress.

It is also noted that meditation practice and physical activity were slightly higher in the low stress group. Screen time, work hours, and alcohol intake were highest in the low stress group which points that this may indicate compensatory mechanisms like social support, flexibility, or better coping strategies despite longer work exposure.

TABLE II: CATEGORYWISE DETAILS - 1

Stress Level	Age	Gender	Occup.	Marital Status	Sleep Duration
0	41.485050	0.594684	78.265781	1.225914	6.164784
1	36.123457	0.438272	96.425926	1.506173	6.671605
2	37.809677	0.448387	90.322581	1.461290	6.332903

TABLE III: CATEGORYWISE DETAILS - 2

Stress Level	Sleep Quality	Physical Activity	Screen Time	Caffeine Intake
0	3.878738	3.435216	4.599668	2.292359
1	3.880247	2.570988	3.577160	1.302469
2	3.801613	2.750000	3.901613	1.629032

TABLE IV: CATEGORYWISE DETAILS - 3

Stress Level	Alcohol Intake	Smoking Habit	Work Hours	Travel Time
0	1.222591	0.730897	8.647841	3.403654
1	0.524691	0.370370	8.080247	2.290123
2	0.754839	0.435484	7.974194	2.625806

TABLE V: CATEGORYWISE DETAILS - 4

Stress Level	Social Interactions	Meditation Practice	Exercise Type
0	3.375415	0.820598	3.342193
1	3.191358	0.530864	3.691358
2	3.025806	0.477419	3.464516

TABLE VI: CATEGORYWISE DETAILS - 5

Stress Level	Blood Pressure	Cholesterol Level	Blood Sugar Level
0	145.348837	232.956811	119.335548
1	128.024691	206.419753	103.395062
2	135.935484	216.596774	108.790323

VII. CONCLUSION

The model reveals that stress is not random but it is greatly influenced by a combination of demographic factors (occupation, age) and physiological health (cholesterol). These insights strengthen the need for systems which can continuously monitor these indicators and alert users to early signs of stress. This will guide them toward preventive or corrective action. The relatively lower performance for medium stress suggests the need for more training data, better feature engineering or use of more advanced models.

REFERENCES

[1] S. A. Han, M. Naqi, S. Kim, and J. H. Kim, "All-day wearable health monitoring system," *EcoMat*, vol. 4, no. 4, p. e12198, 2022.
[2] J. Nurmi and E. S. Lohan, "Systematic review on machine-learning algorithms used in wearable-based ehealth data analysis," *IEEE Access*, vol. 9, pp. 112221–112235, 2021.
[3] I. Ahmad, Z. Asghar, T. Kumar, G. Li, A. Manzoor, K. Mikhaylov, S. A. Shah, M. Hoyhtya, J. Reponen, J. Huusko, *et al.*, "Emerging technologies for next generation remote health care and assisted living," *IEEE Access*, vol. 10, pp. 56094–56132, 2022.
[4] S. Barua, S. Begum, and M. U. Ahmed, "Supervised machine learning algorithms to diagnose stress for vehicle drivers based on physiological sensor signals," in *pHealth 2015*. IOS Press, 2015, pp. 241–248.
[5] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2020, pp. 51–57.
[6] S. Ishaque, N. Khan, and S. Krishnan, "Physiological signal analysis and stress classification from vr simulations using decision tree methods," *Bioengineering*, vol. 10, no. 7, p. 766, 2023.
[7] Y. Haque, R. S. Zawad, C. S. A. Rony, H. Al Banna, T. Ghosh, M. S. Kaiser, and M. Mahmud, "State-of-the-art of stress prediction from heart rate variability using artificial intelligence," *Cognitive Computation*, vol. 16, no. 2, pp. 455–481, 2024.
[8] A. Gobbetti, M. Gjoreski, H. Gjoreski, N. Lane, and M. Langheinrich, "Fedmma-har: Federated learning for human activity recognition with missing modalities using head-worn wearables," *IEEE Pervasive Computing*, 2024.
[9] A. Alsadoon, G. Al-Naymat, and O. D. Jerew, "An architectural framework of elderly healthcare monitoring and tracking through wearable sensor technologies," *Multimedia Tools and Applications*, vol. 83, no. 26, pp. 67825–67870, 2024.
[10] M. A. Miah, E. Rozario, F. B. Khair, M. K. Ahmed, M. M. R. Bhuiyan, and M. Manik, "Harnessing wearable health data and deep learning algorithms for real-time cardiovascular disease monitoring and prevention," *Nanotechnol. Perceptions*, vol. 15, no. 3, pp. 326–349, 2019.
[11] D. C. Peruzzo, B. B. Benatti, G. M. Ambrosano, G. R. Nogueira-Filho, E. A. Sallum, M. Z. Casati, and F. H. Nociti Jr, "A systematic review of stress and psychological factors as possible risk factors for periodontal disease," *Journal of Periodontology*, vol. 78, no. 8, pp. 1491–1504, 2007.
[12] K. Shah, R. Kumari, and M. Jain, "Unveiling stress markers: A systematic review investigating psychological stress biomarkers," *Developmental Psychobiology*, vol. 66, no. 5, p. e22490, 2024.
[13] J. B. Madavarapu, S. Nachiyappan, S. Rajarajeswari, N. Anusha, N. Venkatachalam, R. C. B. Madavarapu, and A. Ahilan, "Hot watch: Iot based wearable health monitoring system," *IEEE Sensors Journal*, 2024.
[14] J. Vigneswari and B. Yamini, "Smart health monitoring system using iot sensors in cloud environment," in *2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG)*. IEEE, 2024, pp. 1–5.
[15] G. Kunapuli, *Ensemble Methods for Machine Learning*. Simon and Schuster, 2023.
[16] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. CRC Press, 2025.
[17] A. Parmar, R. Katariya, and V. Patel, "A review on random forest: An ensemble classifier," in *International Conference on Intelligent Data Communication Technologies and Internet of Things*. Springer, 2018, pp. 758–763.
[18] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization," *Machine Learning*, vol. 32, pp. 1–22, 1998.
[19] H. Wang, F. Yang, and Z. Luo, "An experimental study of the intrinsic stability of random forest variable importance measures," *BMC Bioinformatics*, vol. 17, no. 1, p. 60, 2016.
[20] S. Janitza, E. Celik, and A.-L. Boulesteix, "A computationally fast variable importance test for random forests for high-dimensional data," *Advances in Data Analysis and Classification*, vol. 12, no. 4, pp. 885–915, 2018.
[21] R. Iranzad and X. Liu, "A review of random forest-based feature selection methods for data science education and applications," *International Journal of Data Science and Analytics*, pp. 1–15, 2024.
[22] M. Bicego and F. Escolano, "On learning random forests for random forest-clustering," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 3451–3458.
[23] A. Palczewska, J. Palczewski, R. M. Robinson, and D. Neagu, "Interpreting random forest models using a feature contribution method," in *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)*. IEEE, 2013, pp. 112–119.
[24] A. Palczewska, J. Palczewski, R. Marchese Robinson, and D. Neagu, "Interpreting random forest classification models using a feature contribution method," in *Integration of Reusable Systems*. Springer, 2014, pp. 193–218.