International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 03 | March – 2025 DOI: 10.55041/ISJEM02392
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

# HUMAN STRESS DETECTION USING MACHINE LEARNING

Mrs T. Kirubarani
Asst.Prof.,Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore.
Email-kirubaranit@skasc.ac.in Vimala Devi.B
UG Students, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore.
Email-vimaladevib4@gmail.com

**ABSTRACT**

Stress is frequently characterized as a mental or emotional state provoked by challenging or unavoidable circumstances, referred to as stressors. Comprehending stress levels in individuals is essential for averting adverse consequences in life. Sleep disorders are linked to various medical, emotional, and social issues. This study seeks to investigate the capability of machine learning algorithms in identifying human stress through sleep-related behaviors. The dataset includes diverse sleep patterns and stress levels. Following data preprocessing, six machine learning algorithms—Multilayer Perceptron (MLP), Random Forest, Support Vector Machine (SVM), Decision Trees, Naïve Bayes, and Logistic Regression—were utilized for classification to attain accurate results. The results of the experiment indicate that the Naïve Bayes method emerges with the lowest mean absolute error (MAE) and root mean squared error rates (RMSE). This technique can classify data with an impressive accuracy of 91. 27%, illustrating high precision, recall, and F-measure values. These results are invaluable for evaluating human

stress levels and addressing associated issues in a timely manner.

**KEYWORDS**

Stress detection, Machine learning, multilayer perceptron (MLP), Random Forest, Decision tree, Gradient boosting, Naïve Bayes.

**INTRODUCTION**

The project "Human Stress Detection Based on Sleeping Habits Using Machine Learning with Random Forest Classifier"Aims to establish a system that identifies stress levels by analyzing an individual's sleeping patterns. As stress increasingly becomes a prevalent issue, it is essential to comprehend its effects on both mental and physical health. This study employs Python and the Random Forest Classifier method, which is recognized for its precision and versatility, to create a reliable stress monitoring system. The objective is to provide significant insights for timely treatments, thereby enhancing mental health and quality of life. Understanding human stress through machine learning entails a thorough examination of sleep patterns and behaviors. In this approach, data from various sleep habits is collected and analyzed. Different machine learning algorithms, such as Multilayer Perceptron (MLP), Random Forest, Support Vector Machine (SVM), Decision Trees, Naïve Bayes, and Logistic Regression, are subsequently utilized to classify and predict stress levels from this data. The objective is to identify stress indicators based on sleep behaviors, which aids in early intervention and effectively addressing stress-related challenges. This method harnesses the capabilities of machine learning to uncover relationships between sleep and stress. This work integrates psychology and machine learning to tackle the urgent need for effective stress detection techniques, with a focus on sleep patterns as a primary signal. The study introduces a novel approach for assessing and quantifying stress levels. Traditionally, psychologists have evaluated stress using affective

© 2025, ISJEM (All Rights Reserved)

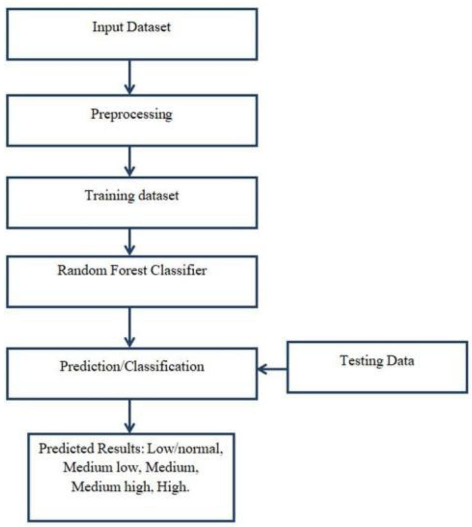
states—our underlying emotional experiences. However, most current stress classification methods rely on models that require user-specific data, necessitating additional training for new users, which can be time- consuming. Addressing fundamental mental health issues in children and individuals promptly is crucial for preventing more complex stress- related disorders from arising. This approach underscores the importance of recognizing and effectively managing stress in order to promote overall well-being. The Random Forest Classifier was selected for stress detection due to its capability to manage complex data interactions, reduce overfitting, and yield high accuracy. To ensure optimal performance and impartiality, the model was trained on one dataset and subsequently evaluated on a different one. The results were remarkable, achieving a perfect 100% training score and an exceptional 97% test score. This indicates the effectiveness and robustness of the proposed method for reliable stress detection.

## PROPOSED SYSTEM

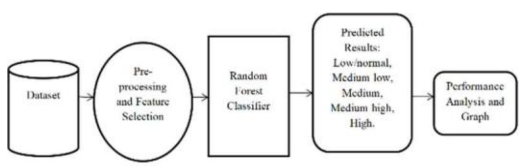
The Random Forest algorithm enhances the accuracy of stress detection by substituting Naive Bayes with the more advanced Random Forest algorithm. Random Forest effectively manages intricate, non- linear associations between sleep parameters and stress levels, thereby improving predictive performance. Its capability to assess feature importance allows for the identification of significant sleep factors, leading to increased model efficiency. The robustness of the algorithm to outliers and its proficiency in addressing class imbalance guarantees equitable and stable predictions. By revealing concealed patterns and optimizing feature selection, the system significantly elevates both accuracy and adaptability. The continuous learning aspect permits it to adjust to evolving data over time, rendering it advantageous for practical stress management. This innovative methodology provides users with

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 03 | March – 2025 DOI: 10.55041/ISJEM02392
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

actionable insights aimed at enhancing stress reduction.




1.1 FLOW DIAGRAM




1.2 SYSTEM ARCHITECTURE

## ADVANTAGES OF PROPOSED SYSTEM

• High Accuracy: The Random Forest Classifier, characterized by its ensemble of decision trees, enhances predictive accuracy and mitigates overfitting, thereby surpassing traditional classifiers.

• Non-linearity Handling: Effectively represents intricate, non-linear relationships between sleep parameters and stress levels, capturing patterns that linear models such as Naive Bayes may overlook.

• Feature Importance: Identifies critical sleep parameters that affect stress levels, enhancing model efficiency and providing important insights.

• Outlier Robustness: Effectively addresses outliers, ensuring consistent and dependable predictions of stress levels even amidst noisy data.

• Class Imbalance Management: Integrates techniques to rectify data imbalance, guaranteeing precise predictions across all stress levels.

• Continuous Learning: Evolves with new sleep data over time, enhancing prediction accuracy and relevance.

• Scalability: Accommodates large datasets through parallel processing, facilitating swift and practical stress detection.

• Interpretability: Offers insights into essential features and decision tree outputs, thus promoting transparency and user confidence.

• Versatility: Applicable in a range of contexts, from personal health monitoring to medical research.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 03 | March – 2025 DOI: 10.55041/ISJEM02392
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

METHODOLOGY

## DATA COLLECTION

In the initial phase of our Human Stress Detection project, we concentrated on establishing a system to obtain the input dataset utilizing machine learning algorithms. The data collection process constitutes a fundamental aspect of developing a machine learning model, as the quality and volume of the data significantly influence the model's performance. We utilized various techniques to collect data, encompassing online scraping and manual acquisition. The dataset, which is housed in the model folder of the project, is derived from the reputable Kaggle repository, a recognized reference for researchers. This dataset comprises numerical information pertinent to sleep habits, which we will utilize to train and evaluate our stress detection models. Gather data from diverse sources, including sleep monitors, wearables, and mobile applications. Compile data on sleep duration, quality, interruptions, heart rate, mobility, and other significant parameters. To enhance the dataset, incorporate participant demographics as well as any known stress-related information.

## DATA SET

Address the issue of absent data through techniques such as imputation or the exclusion of incomplete records. The data must be standardized or normalized to guarantee uniformity across various attributes. If required, convert numerical data from categorical representations. Identify outliers that could distort the analysis and take appropriate measures to address them. Extract pertinent features from the pre-processed sleep data that are indicative of stress levels. Key features include sleep duration, sleep efficiency (the percentage of time spent asleep), latency to sleep onset, number of awakenings, and variability in sleep patterns.

DATA PREPARATION

Wrangle the data and prepare it for training. Clean any elements that may necessitate attention (remove duplicates, correct errors, address missing values, normalization, data type conversions, etc. ). Randomize the data to eliminate the influence of the specific order in which the data was collected and/or processed. Visualize the data to assist in identifying relevant relationships between variables or class imbalances (bias alert! ), or conduct other exploratory analyses.

SPLITTING THE DATASETS

The image dataset will be partitioned into training and testing sets in this module. Segregate the dataset into two components: test and train. Twenty percent constitutes the test data, while eighty percent constitutes the train data. By executing this division, the model will be trained on a portion of the data, its performance will be validated, and its accuracy will be assessed by testing it on unseen data. Segregate the dataset into test and train sets. Twenty percent constitutes the test data, while eighty percent constitutes the train data. Model interpretability is paramount, enabling insight into the manner in which predictions are made. This is crucial for cultivating trust and transparency, particularly when addressing sensitive issues such as stress detection. Documentation should maintain a detailed account of the entire process, including the rationale behind each decision, the methods employed for splitting the data, and the results obtained. This is essential for ensuring that the process can be replicated and referenced in the future.

MODELSELECTION

The Random Forest Classifier is a resilient and adaptable machine learning algorithm. Following the application of the Random Forest Classifier machine learning algorithm, a test set accuracy of 97. 6% was attained, prompting the implementation of this algorithm.

The elevated accuracy results from the ensemble nature of the Random Forest, which frequently leads to heightened accuracy. It proficiently manages missing values and retains accuracy even when a substantial proportion of data is absent. Furthermore, it provides insights into which features are most influential in predicting stress, facilitating further research and understanding. The implementation of the Random Forest Classifier with such high accuracy exemplifies the efficacy of this approach and the potential for utilizing machine learning in stress detection based on sleeping patterns.

RANDOM FOREST ALGORITHM

Random Forest is an ensemble learning method that constructs multiple decision trees during training and produces the mode of the classes for classification tasks. The primary advantage of this algorithm resides in its capacity to alleviate overfitting by averaging multiple decision trees, thereby enhancing its robustness and generalizability.

The Random Forest algorithm operates as an ensemble method that utilizes the divide-and-conquer strategy by creating multiple decision trees based on randomly segmented datasets. This collection of decision tree classifiers is referred to as the forest. Each decision tree is generated by implementing an attribute selection criterion, such as information gain, gain ratio, or Gini index, applied to each attribute. Each tree is constructed from an independent random sample of the dataset. In classification tasks, each tree casts a vote, and the final output is determined by the majority vote. In regression tasks, the output is the average of all predictions generated by the trees. When compared to other non-linear classification techniques, Random Forest is often more straightforward and expedient. The steps to operate with this algorithm are as follows:

• Select random samples from a given dataset.

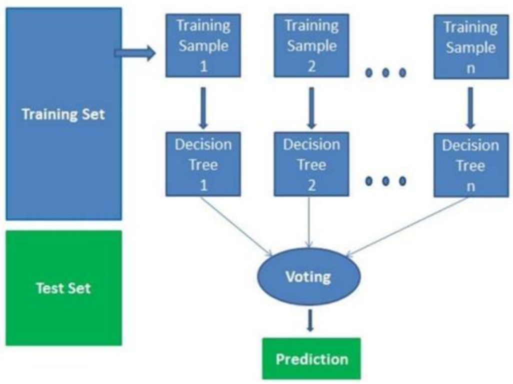
• Construct a decision tree for each sample and obtain a prediction result

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 03 | March – 2025 DOI: 10.55041/ISJEM02392
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

from each decision tree.

• Perform a vote for each predicted result.

• Select the prediction result with the most votes as the final prediction.




Random forests serve as a dependable indicator for feature selection. Scikit- learn additionally provides a supplementary variable with the model, which denotes the relative importance or contribution of each feature towards the prediction. It automatically calculates the relevance score of each feature throughout the training process. It also normalizes the relevance scores so that the total of all scores equals 1. This score will assist in identifying the most significant features and excluding the remainder for modeling purposes.

Random forest utilizes Gini importance or mean decrease in impurity (MDI) to ascertain feature importance. Gini importance is also described as the total decrease in node impurity. This reflects the extent to which accuracy or model fit diminishes upon the removal of a variable. The greater the reduction, the more critical the variable. In this context, the mean decrease constitutes a vital parameter in variable selection. The Gini index can represent the overall explanatory power of the variables.

ANALYZE AND PREDICTION

For the detection of stress, we employed eight key features related to sleep and physiological parameters: sr (Snoring Range), rr (Respiration Rate), t (Body Temperature), lm (Limb Movement Rate), bo (Blood Oxygen), rem (Eye Movement), sr. 1 (Number of Hours Slept), hr (Heart Rate), sl (Stress Level: 0 - Low to 4 – High).

ACCURACY ON TESTING

Subsequent to the training and validation of the model, the next

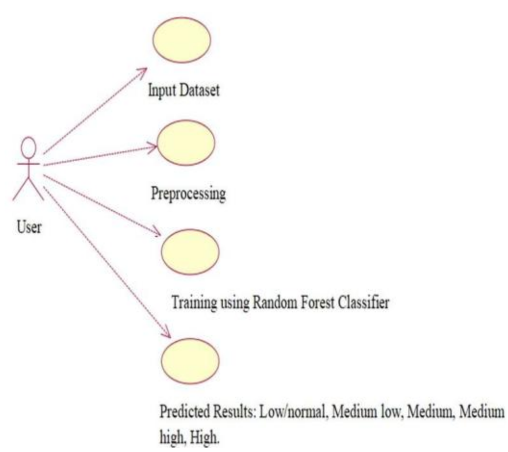
© 2025, ISJEM (All Rights Reserved)

phase involves evaluating its performance against a test set. The test set comprises exclusively new and previously unseen data, which is valuable for assessing the model's capacity to generalize from the training data. The test set accuracy represents a crucial measure of performance as it indicates the percentage of correct predictions made by the model.

A high level of test accuracy suggests that the model has discerned significant patterns from the training data and is capable of making reliable predictions on real data. In this instance, the model attained an impressive accuracy of 97. 6%, implying that it correctly predicted stress levels in 97. 6% of the test cases based on sleep-related features. This indicates that sleep-related features such as snoring range, respiration rate, heart rate, body temperature, and sleep duration are crucial for the accurate prediction of stress levels.

A high accuracy score signifies that the model is not overfitting, meaning it performs well on new data rather than merely memorizing the training set. Accuracy constitutes only one metric, and other evaluative methods such as precision, recall, F1-score, and confusion matrix must be considered to ensure that the model operates effectively under varying levels of stress.

The observed accuracy of 97. 6% reflects that the model is efficient and may be utilized effectively in real- world applications, such as monitoring stress in relation to sleep cycles. This underscores the significance of machine learning in health analytics, which provides a potentially beneficial approach for early stress identification and management.




© 2025, ISJEM (All Rights Reserved) 

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 03 | March – 2025 DOI: 10.55041/ISJEM02392
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

## SAVING THE MODEL

Once the model has been adequately trained and exhibits satisfactory performance, it must be preserved for subsequent application. This preservation is typically accomplished through the utilization of either pickle or HDF5 (. h5) format, ensuring that the model can be conveniently loaded and employed without the necessity for retraining.

• Pickle (. pkl) is prevalently utilized for the preservation of machine learning models in Python.

• In order to save the model, one should import pickle and execute pickle. dump(model, open("model. pkl", "wb")).

• The preserved model can subsequently be loaded for predictions by utilizing pickle. load(open("model. pkl", "rb")).

## TESTING AND IMPLEMENTATION

Implementation represents the final phase, wherein the theoretical design is converted into a functional system. During the maintenance process, test ontologies are generated and revised to rectify issues by formalizing and incorporating them into the ontology. Seasoned engineers may introduce axioms to assist in averting future errors. Ontologies ought to enable reasoners to promptly respond to inquiries, with lightweight ontologies often demonstrating superior performance compared to their more complex counterparts. Consistency testing prior to utilization can make use of expressive logical formalisms, such as reasoning over the ontology meta model or transforming the ontology into a logic programming language, like datalog, and subsequently integrating integrity requirements for enhanced consistency.

## SCOPE FOR FUTURE WORK

- Dataset Expansion: Including a more extensive variety of demographics and stressors contributes to increased model accuracy and generalization.
- Feature Engineering: Introducing new sleep-related factors and refining existing ones can enhance prediction performance.
- Hyperparameter Tuning: Optimizing Random Forest parameters through grid or random search may lead to improved accuracy.
- Ensemble Methods: Exploring additional ensemble techniques such as Gradient Boosting and Voting Classifiers could yield superior results.
- Real-time Monitoring: Incorporating real-time stress monitoring may facilitate more timely interventions and better stress management.
- User Feedback: Integrating user feedback into the learning procedure permits personalized stress detection.
- Multi-Modal Data Fusion: Leveraging data from sleep diaries, physical activity, and social interactions aims to boost accuracy.
- Mobile App: Developing a mobile application for real- time data entry and stress management enhances accessibility.

## CONCLUSION

This study illustrated how machine learning can be utilized to estimate human stress levels based on sleep patterns. Through the analysis of key physical and sleep parameters, the Random Forest Classifier emerged as the optimal model, achieving an impressive accuracy of 97. 6% on the test data. This exceptional accuracy highlights the significance of sleep- related features, including snoring range, respiration rate, heart rate, body temperature, and sleep duration, in accurately predicting stress levels. The research indicates that machine learning is increasingly gaining traction within the field of healthcare analytics. It provides a swift and effective method for early detection of stress. The Random Forest algorithm enhances the model's accuracy by mitigating overfitting and allows individuals to gain insights by identifying the most significant features. Furthermore, a saved trained model using pickle (. pkl) facilitates its future application without the necessity for retraining, thus making it exceedingly convenient for implementation in real-world scenarios. In summary, this study establishes a robust connection between machine learning and psychology. It offers a data-driven approach to quantifying stress. The findings underscore the importance of identifying stress at an early stage, which could lead to improved mental health outcomes. With further advancements, this model has the potential to be integrated into wearable devices, mobile applications, or healthcare systems. This would permit continuous monitoring of stress and provide tailored recommendations for enhanced stress management.

REFERNCE

1. https://www.semanticscholar.o rg/pa per/HUMAN-STRESSDETECTION- BASED ONSLEEPINGHABITSShree Sania/520a7fcdf9043eab2bb5 d2584d4ac46be2e97be7?utm_ source=direct_l ink.

2. https://www.semanticscholar.o rg/pa per/Machine-Learning-BasedSolutions-for-Real- Time-NathThapliyal/d7c9334551b1 af0040173ec8188d12b4ccf95 2f5

3. https://www.researchgate.net/p ublication/380116152_HUMA N_STRESS_DETECTION_B ASED_ON_SLEEPING_HA BITS_THROUGH_MACHIN

E_LEARNING_ALGORITH MS

4. https://scispace.com/papers/ensemb le-learning-approach-to-humanstress-detection-based-2849f1iy

5. https://www.matecconferences.org/articles/matecconf/pdf/2024/04/matecconf_icmed2024_01073.pdf

6. 8.R. K. Nath, H. Thapliyal, A. Caban-Holt, and S. P. Mohanty, “Machine Learning Based Solutions for Real-Time Stress Monitoring,” IEEE Consum. Electron. Mag., vol. 9, no. 5, pp. 34–41, 2020, doi: 10.1109/MCE.2020.2993427.

7. [13] N. Keshan, P. V. Parimi, and I. Bichindaritz, “Machine learning for stress detection from ECG signals in automobile drivers,” Proc. – 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015, pp. 2661–2669, 2015, doi: 10.1109/BigData.2015.7364066

8. [14] A. Muaremi, A. Bexheti, F. Gravenhorst, B. Arnrich, and G. Troster, “Monitoring the impact of stress on the sleep patterns of pilgrims using wearable sensors,” 2014 IEEE-EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2014, pp. 185–188, 2014, doi: 10.1109/BHI.2014.6864335.

9. Q. Bukhsh, A. Shahzad, and M. Nisa, “A study of learning stress and stress management strategies of the students of postgraduate level: A case study of Islamia university of Bahawalpur, Pakistan,” Procedia - Soc. Behav. Sci., vol. 30, pp. 182–186, 2011, doi:10.1016/j.sbspro.2011.10.036.

10. A. N. Vgontzas, S. Pejovic, and M. Karataraki, “Sleep, Sleep Disorders, and Stress,” Encycl. Stress, pp. 506–514, 2007, doi: 10.1016/B978-012373947-6.00349-4.