iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.83020

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

SleepSynergetic Stress Predictor: A Data-Driven Approach to Anticipating Stress Levels During Sleep

Anjali Choudhary
Dept. of Information Technology Delhi Technological University Delhi, India

Abstract— This study paper has explored the complex relationship between stress and sleep, defining stress as a condition of mental or emotional pressure influenced by many stimuli. Our focus lies in predicting stress levels during sleep using a comprehensive Stress Behavior Detection Mechanism. The mechanism employs real-time physiological signal monitoring, as well as demographic factors have been also taken into consideration. Notably, 'snoring rate' followed by 'sleep_hours' and 'eye_movement are identified as pivotal indicators of stress levels during sleep, while 'limb_movement,' and 'heart_rate' along with 'age' also contribute significantly. The model encompasses various classifiers, including Logistic Regression, Decision Tree, Random Forest, SVM, KNN, and Gaussian Naive Bayes. The stress levels are classified into five categories: "Low/Normal," "Medium Low," "Medium," "Medium High," and "High." The model's training involves a detailed analysis using confusion matrices and classification reports for each classifier. Machine learning models have been used to calculate F1-score, recall, and accuracy. In our model, random forest has outperformed all the available models having an accuracy of 97.6%. Lastly, we also discuss future directions and possible research challenges.

Keywords— Stress, Sleep, Stress Prediction, Random Forest, KNN, SVM, Decision Tree, Logistic Regression, Gaussian Naïve Bayes

I. INTRODUCTION

The physiological process of sleep is essential for preserving one's physical, mental, and emotional health. The body repairs itself, integrates memories, and modifies a number of internal processes while we sleep. The intricate and substantial relationship between stress and sleep is attributed to the profound impact that stress can have on both general health and the quality of sleep.

Stress has a profound effect on sleep architecture, leading to disruptions in sleep patterns, sleep fragmentation, and alterations in sleep stages. Moreover, stress during sleep can trigger physiological responses, including elevated heart rate, increased cortisol levels, and changes in respiration and body temperature. These reactions have far-reaching effects on long-term health, including an increased risk of cardiovascular illnesses, mental health issues, and weakened immune systems, in addition to disrupting sleep. Understanding stress levels during sleep is critical for comprehensive health management. Accurate assessment and prediction of stress during sleep can facilitate targeted interventions to improve sleep quality, mitigate health risks associated with stress, and enhance overall well-being. Prolonged stress during sleep is linked to a higher risk of hormone imbalances, weakened immune systems, mental health conditions like anxiety and depression, and cardiovascular illnesses. Prolonged exposure to stress during sleep can exacerbate these conditions. Continuous stress during sleep can lead to daytime fatigue, cognitive impairment, mood disturbances, and decreased productivity, impacting an individual's overall quality of life and well-being. By incorporating advanced data-driven approaches and physiological parameters, this project aims to enhance health monitoring capabilities. Real-time analysis and predictive models offer opportunities for personalized health management strategies, thereby fostering improved sleep quality and overall well-being.

Accurate assessment and prediction of stress during sleep facilitate targeted interventions. Tailored strategies can be implemented to alleviate stress-related sleep disturbances, enhance sleep quality, and mitigate associated health risks.

In this discipline, a wide range of technologies are employed, such as Doppler radar, radio waves, and smartphones. The gathered data is analyzed and assessed using data mining and machine learning techniques to extract important sleep metrics and indicators. With home-based monitoring, one may assess and manage their daily sleep quality while relaxing in their own house. In this discipline, devices including Doppler radars, radio waves, and smartwatches are all employed. Data mining and machine learning techniques are applied to assess and analyse the gathered data by extracting relevant sleep metrics and markers.

For individuals and researchers attempting to gain a better understanding of the connections between, for example, their daily and evening routines, sensor-monitored sleep patterns at home may be beneficial.

Thorough analysis and classification of our study unveiled the pivotal role of specific factors in predicting stress levels during sleep. Notably, the snoring rate emerged as a predominant predictor, signifying its substantial influence on stress levels. Additionally, variables such as sleeping hours, eye movement, and heart rate demonstrated significant contributions, underscoring their relevance in accurately predicting stress levels during sleep.

"SleepSynergetic Stress Predictor" endeavors to unravel the complexities of stress and its impact on sleep using a robust data-driven approach. By integrating various physiological parameters and leveraging advanced analytical methodologies, this project aims not only to predict stress levels during sleep but also to pave the way for tailored interventions aimed at enhancing sleep quality and fostering holistic well-being.

II. RELATED PRIOR WORKS AND RESEARCH GAP

A. Related Prior Studies

the investigation of personalized prediction using self-supervised learning on multimodal time-series data for recurrent stressful occurrences is elucidated in [1]. Stress and stress-related mental problems are the subject of a large number of studies in the fields of machine learning, deep learning, and data preparation techniques., as exemplified in [2]. Additionally, research on fatigue life prediction, taking into account mean stress effects based on random forest methodology, is documented in [3]. An innovative application of IoT systems and sensors for detecting the position of a reclining person is also illustrated in [4].

Various wearable devices play a pivotal role in stress detection and monitoring, as evidenced by studies in [1] and [5], employing diverse principles such as the Triangular Principle in a Wearable Sensor System as described in [5]. Additionally, non-wearable solutions are available for sleep regulation [6]. As suggested in [5], virtual reality technology is becoming a useful tool for stress monitoring. A succinct summary of the studies on stress detection is provided in [7]. Additionally, a study [8] looks into how stress affects a person's ability to communicate across muscles.

One foundational study [9], forming the basis for our research, delves into stress level prediction through sleep data using various physiological factors. While existing smart sleep study techniques and approaches are evident in [10], and [11], our research goes beyond by incorporating multiple classification models for a more efficient solution. Additionally, we explore demographic factors, such as age, to enhance the comprehensiveness of our investigation.

B. Research Gap

The following are unaddressed issues from prior research that offer opportunities for enhancement, paving the way for the implementation of a superior and more reliable model:

- Users lack awareness regarding the significance of achieving quality sleep and its direct impact on stress levels.
- A comprehensive automated system for stress detection is currently absent, leaving users without an efficient means of identifying and managing stress.
- The reliance on a restricted set of physiological parameters has limited the scope of stress detection capabilities.
- The utilization of a solitary database introduces potential constraints on the diversity of sleep patterns and demographic representations within the research.
- While the predictive models employed in the initial research demonstrated commendable accuracy, there is ample room for enhancement.
- Previous efforts were predominantly fixated on a narrow range of classifiers, lacking the desired diversity in approach.
- Expanding the repertoire of machine learning models is essential, allowing for a nuanced understanding of their individual strengths and weaknesses in the context of predicting stress levels during sleep.
- Enhancements to the system can ensure robustness and security, safeguarding user data storage and privacy.
- Incorporating deep learning techniques can further elevate the reliability and precision of the results, providing users with more accurate insights into their stress levels during sleep.

III. STUDY AND PROPOSED METHODS

To discern stress levels across a spectrum (ranging from 0 to 4) using the aggregated sleep data, we have implemented a diverse set of conventional machine learning models, including KNN, SVM, Decision Tree, Random Forest, Logistic Regression, and Gaussian Naive Bayes.

The deliberate integration of multiple classifier models serves the primary purpose of showcasing the superiority of our proposed model. To facilitate a comprehensive understanding, a comparative analysis of various evaluation scores is presented, encompassing metrics such as Recall, Precision, F1-Score, and Accuracy. A confusion matrix is also used as a visual aid to improve understanding and offer a more detailed evaluation of the model's performance.

A. Decision Tree Classifier

DTs are a popular, practical, and fast method for mining and finding patterns in big, complicated datasets. Because it creates the foundation for information extraction and modelling from massive databases, this topic is essential. Experts and investigators alike look for novel approaches to expedite, streamline, and improve the process. Among the applications for DTs include data mining, machine learning, information extraction, text mining, and pattern recognition. [12, 13]. A decision tree is composed of decision nodes and leaf nodes. Every decision node includes many branches, each of which addresses the result of the test X, and corresponds to a test X over a particular attribute of the input data. A class that is the outcome of a case judgement is represented by each leaf node. [14]. Once each row of data is analysed, connection weights are adjusted in deep learning models based on the degree of error in the output relative to the expected result.

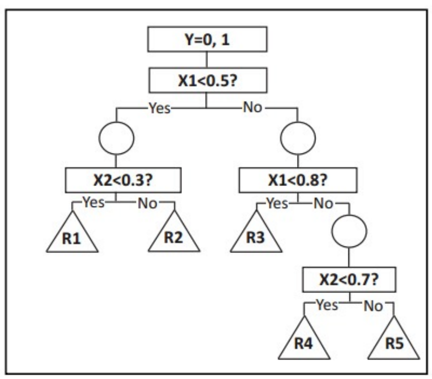


Fig 1. An example architecture of DT Classifier [15]

B. Support Vector Classifier

Producing a separation surface that maximizes for the classes in the training set is the main objective of support vector machines. To reiterate, SVM makes it possible to maximize a model's capacity for generalization. The Structural Risk minimizing (SRM) concept permits the minimizing of a bound on a model's generalization error, in contrast to the conventional philosophy of empirical risk minimization approaches, which seeks to minimize the mean squared error on the set of training data. For our training purposes, take into consideration the following two-class issue dataset: (x1, y1).... (xN, yN), where yi □ +1,−1 indicates the results from the sample, etc., and xi(□ R D) is a feature vector from the training data that represents the ith sample. The main goal is to find a decision function that can correctly predict y from x. A non-linear Support Vector Machine Classifier is used to conclude. f(x) = sign(g(x)) in the input vector, for some x

$$g(x) = \sum_{n=1}^{l} \omega_i * K(x, z_i) + b \quad (1)$$

In this case, x is a member of the class if f(x) = +1, and it is not a member if f(x) = −1. The number of support vectors is indicated by l. Each training example is indicated by zis. The kernel used to map data to higher dimensions is called K. Conventional kernels use dot products: K(x,z) equals k(x.z). A degree d polynomial kernel expressed as k(x) = (1 + x)^d [28]

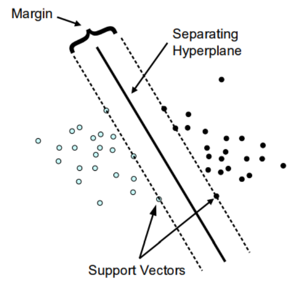


Fig 2. An example of SVM [16]

C. Random Forest Classifier

To determine how to categorise the input vector, the random forest classifier employs a massive number of trees, each of which is constructed using a randomly selected vector unrelated to the input vector [17]. In this work, a random forest classifier was used, which builds a tree with a distinct random feature or feature combination used by each node. We employed bagging, which is the process of randomly choosing N samples from the original training set and replacing them with new ones, to build a training dataset for each feature/feature combination [18]. We use the tree predictors of the forest to determine which label is most common for each data point [17]. Choosing a pruning strategy and an attribute selection metric was essential to building a decision tree. The process of choosing characteristics for decision tree induction usually entails giving the attribute a quality metric. Two of the most used attribute selection metrics in decision tree induction are the Information Gain Ratio criteria and the Gini Index [19]. The random forest classifier uses the Gini Index, which gauges an attribute's purity in relation to the classes. Furthermore, it outperformed every other classifier that was in use.

$$\sum\sum_{j\neq i} (f(C_i ,T)/|T|)(f(C_f ,T)/|T|) \qquad (2)$$

here $f(C_i, T)/|T|$ is the probability that the sample belongs to class Ci.

D. Logistic Regression Classification

Despite its name, Logistic Regression is used for classification, not regression. It models the relationship between a dependent binary variable (in the traditional sense) and one or more independent variables. It estimates the probability that a given input belongs to a particular class similar to other models, Logistic Regression requires the selection of relevant features, which might include physiological, behavioral, and psychological factors that could influence stress levels.

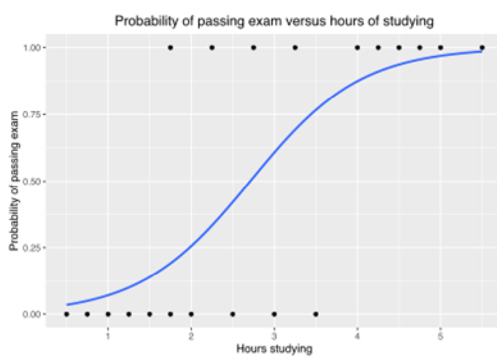


Fig 3. An example of Linear Regression [20]

E. KNN Classification

KNN is a simple yet effective algorithm used for both classification and regression tasks. In this case, it would be applied for regression to predict stress levels on a scale of 0 - 4. To predict stress levels, various features could be considered. These might include physiological measurements, behavioral data, and psychological factors.

Fig 4. An example of a K-Nearest Neighbor [21]

When a new individual's data is provided, the KNN algorithm looks for the K most similar individuals (neighbors) from the historical dataset based on the selected features. Similarity is determined by calculating the distance (using a chosen distance metric like Euclidean or Manhattan distance) between the new data point and the existing data points. The K nearest neighbors are identified, and their known stress levels (which are on a scale of 1-5) are considered.

F. Evaluation Metrics

When doing our studies, we mainly relied on the confusion matrix, which shows the percentages of positive, negative, and false positive outcomes. To assess the effectiveness of our models we used a variety of measures to assess our models, such as the F-score, the precision of our predictions, and a harmonic average of precision and recall. Precision quantifies the likelihood that a prediction will come true, whereas recall quantifies the likelihood that a positive sample will be found.

While accuracy is often used as a KPI, it isn't always the best approach to measure a categorization model's effectiveness. Some situations may benefit more from using alternate measurements such as precision, recall, or F1-score.

$$Accuracy = \frac{T.P + T.N}{T.P + T.N + F.P + F.N} \quad (3)$$

A model's accuracy is proportional to the number of its accurate predictions per one incorrect one. One alternative definition of accuracy is a model's ability to reliably predict the positive class with a minimum of false positives.

$$Precision = \frac{T.P}{T.P + F.P} \quad (4)$$

Recall in machine learning is the percentage of true positive samples that were accurately predicted in a dataset. The accuracy with which a model recognizes positive examples while simultaneously minimizing the amount of false negatives it generates is quantified by the recall.

$$Recall/TruePositiveRate = \frac{T.P}{T.P + F.N} \quad (5)$$

False positive rate is a metric that measures the proportion of negative instances in a dataset that are incorrectly classified as positive by a model. In other words, a false positive rate is the probability of a negative instance being classified as positive.

$$FalsePositiveRate = \frac{F.P}{F.P + T.N} \quad (6)$$

The F1 score is a well-liked evaluation metric in machine learning, as it provides a succinct value that balances precision and recall to characterize a model's performance in a binary classification challenge. The harmonic mean of recall and accuracy scores is used to calculate the F1 score.

$$F1 = 2.\frac{Precision.Recall}{Precision + Recall} \quad (7)$$

IV. NOVAL CONTRIBUTIONS

The novel contributions of SleepSynergetic are:

- Addition of demographic factor—age—beyond physiological factors for a holistic view of stress during sleep.
- Inclusion of age provides comprehensive insights into stress dynamics during sleep, enhancing analysis depth.
- Utilization of multiple databases increases diversity in sleep patterns and demographic representation.
- Advanced feature importance analysis identifies dominant predictors, refining predictive model accuracy and reliability.
- Introduction of new classification models and accuracy improvements signify overall enhancement in predictive capabilities.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

V. RESULTS AND DISCUSSIONS

In our study, we applied six models for predicting stress levels while the individual is sleeping. Here is a tabular representation of quantitative values of different evaluation metrics.

Quantitative Performance Evaluation				
Model Name	Accuracy	Precision	Recall	F1 Score
KNN	95.87%	96%	96%	0.96
Decision Tree	96.34%	96%	96%	0.96
Random Forest	97.62%	96%	96%	0.96
GNB	96.66%	97%	97%	0.97
SVM	96.5%	97%	97%	0.97
Logistic Regression	95.7%	96%	96%	0.96

Table 1. Quantitative Performance Evaluation

From the above table we can say that Logistics regression gives the lowest performance compared to other models and Random Forest gives the highest accuracy among all the classifiers.

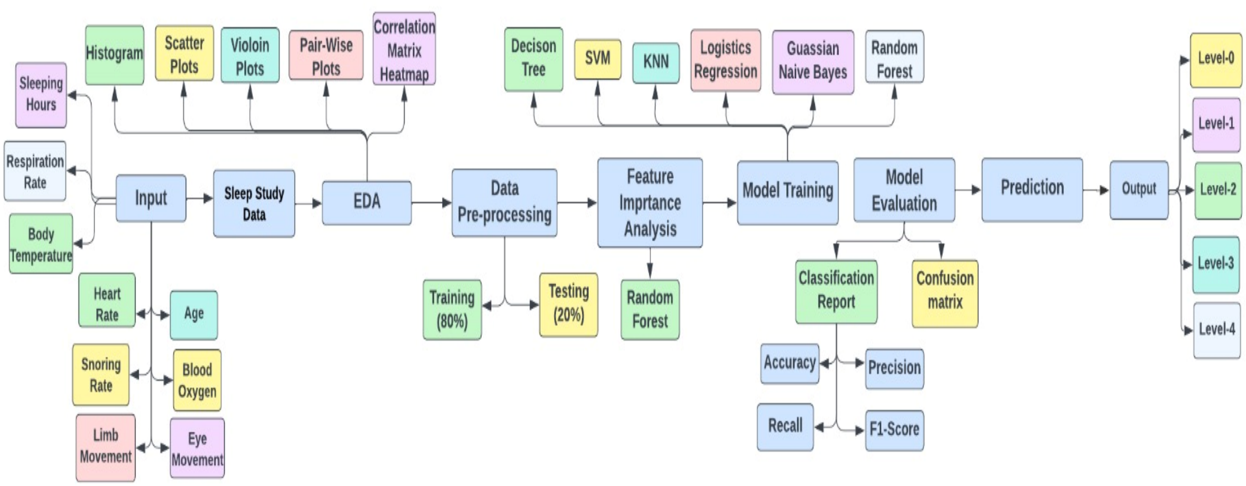


Fig 5. Detailed Proposed Architecture of SleepSynergetic

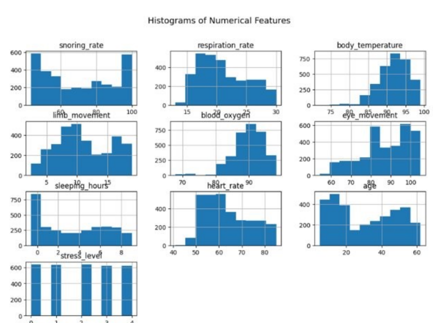


Fig 6. Histograms of Numerical Feature

The above image gives the dependencies of physiological as well as demographic factors wrt to the data points and so a histogram has been made reflecting all the 9 factors.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

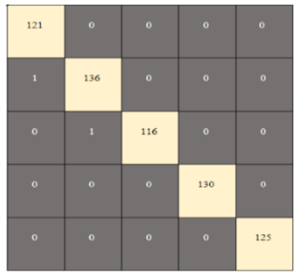


Fig 7. Confusion Matrix for Random Forest

The above image is used to define the performance of a classification algorithm and here the performance of random forest has been shown.

VI. CONCLUSION AND FUTURE WORK

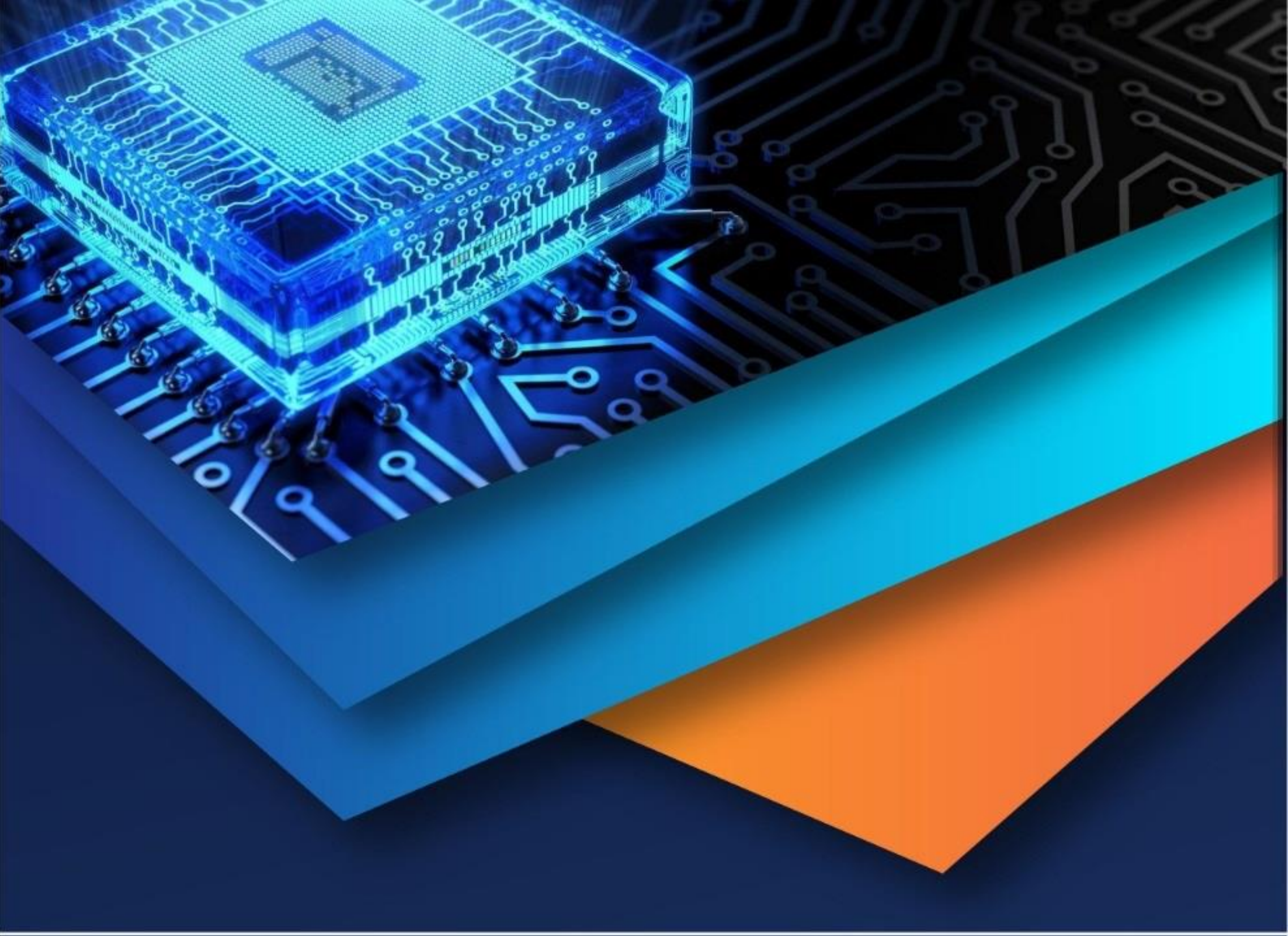
In conclusion, SleepSynergetic's groundbreaking achievements in sleep science set it apart as a transformative force in stress level prediction during sleep. Departing from conventional approaches, the project relies on five distinct datasets, addressing critical limitations and ensuring a robust, universally applicable model. The inclusion of the "Age" factor represents a significant evolution, recognizing the nuanced influence of age on sleep patterns and stress responses, making SleepSynergetic a comprehensive and inclusive model.

The project's meticulous approach is evident in the careful selection and prioritization of impactful features, with the snoring rate emerging as a standout contributor to stress level predictions. The commitment to precision is further demonstrated through the integration of information from diverse datasets and the utilization of a range of classifiers, ensuring versatility and adaptability across various conditions in sleep science. Looking ahead, SleepSynergetic's future prospects include real-time monitoring, personalized recommendations, large-scale studies, expanded feature sets, mobile application development, and longitudinal studies. These initiatives underscore the project's dedication to advancing sleep-related stress management and overall well-being. By empowering individuals to take charge of their sleep health and stress management, SleepSynergetic not only contributes to individual well-being but also lays the foundation for future advancements in understanding the complex relationship between sleep and stress in diverse populations. SleepSynergetic is not just a milestone in sleep science; it is a pioneering contribution that promises continued innovation and deeper insights into the intricacies of sleep patterns and stress responses.

REFERENCES

[1] L. Rachakonda, S. P. Mohanty, E. Kougianos, and P. Sundaravadivel, “Stress-Lysis: A DNN-Integrated Edge Device for Stress Level Detection in the IoMT,” IEEE Trans. Conum. Electron., vol. 65, no. 4, pp. 474– 483, Nov 2019.
[2] L. Rachakonda, S. P. Mohanty, and E. Kougianos, “iLog: An intelligent device for automatic food intake monitoring and stress detection in the IoMT,” IEEE Trans. Conum. Electron., vol. 2, no. 66, pp. 115–124, May 2020.
[3] W. Koczkodaj, J. Masiak, M. Mazurek, D. Strzaka, and P. F. Zabrodskii, “Massive Health Record Breaches Evidenced by the Office for Civil Rights Data.” Iran J Pub. Hea., vol. 48, no. 2, pp. 278–288, 2019.
[4] S. Adedoyin, W. Fernando, A. Aggoun, and K. M. Kondoz, “Motion and Disparity Estimation with Self Adapted Evolutionary Strategy in 3D Video Coding,” IEEE Trans. Conum. Electron., vol. 53, no. 4, pp. 1768–1775, 2007.
[5] B. K. Wiederhold, I. T. Miller, and M. D. Wiederhold, “Using Virtual Reality to Mobilize Health Care: Mobile Virtual Reality Technology for Attenuation of Anxiety and Pain,” in Proc. IEEE Consum. Electron. Mag., vol. 7, no. 1, pp. 106–109, 2018.
[6] F. Sannholm, J. Paalasmaa, and L. Leppakorpi, “System for Determining the Quality of Sleep,” US Patent 20 160 213 309A1, 2015.
[7] 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., pp. 1–1, 2020.
[8] S. Lin, H. Zhang, Y. Gao, M. Du, and M. Vai, “The Effects of Muscle Stress on Signal Transmission in the Intra-Body Communication,” in Proc. IEEE Int. Conf. Consum. Electron. - China (ICCE-China), 2016, pp. 1–3.
[9] N. Schneiderman, G. Ironson, and S. Siegel, “Stress and Health: Psychological, Behavioral, and Biological Determinants.” Annu Rev Clin Psychol, vol. 1, p. 607628, 2005.
[10] D. Kim, H. Han, and R. Park, “Gradient Information-Based Image Quality Metric,” IEEE Trans. Conum. Electron., vol. 56, no. 2, pp. 930– 936, 2010.
[11] T. Kawase, M. Okamoto, T. Fukutomi, and Y. Takahashi, “Speech Enhancement Parameter Adjustment to Maximize Accuracy of Automatic Speech Recognition,” IEEE Trans. Conum. Electron., vol. 66, no. 2, pp. 125–133, 2020.
[12] Oded Z Maimon and Lior Rokach. Data mining with decision trees: theory and applications, volume 81. World scientific, 2014.
[13] Priyanka and Dharmender Kumar. Decision tree classifier: A detailed survey. International Journal of Information and Decision Sciences, 12(3):246–269, 2020.
[14] Quinlan, jr: C4. 5 programs for machine learning, morgan kaufmann, san mateo, california (1992).
[15] LU, Y. (2015). Decision tree methods: Applications for classification and prediction. Shanghai Archives of Psychiatry, 27(2), 130-135.

[16] Kim, H., Pang, S., Je, H., Kim, D., & Yang Bang, S. (2003). Constructing support vector machine ensemble. Pattern Recognition, 36(12), 2757-2767.
[17] Leo Breiman. Random forests–random features. technical report 567, statistics department, university of california, berkeley,. 1999.
[18] Leo Breiman. Bagging predictors. Machine learning, 24(2):123–140, 1996.
[19] J Ross Quinlan. C4. 5: programs for machine learning. Elsevier, 2014.
[20] Logistic regression. (2023, December 19). In Wikipedia. https://en.wikipedia.org/wiki/Logistic_regression
[21] Zhang, M., & Zhou, Z. (2007). ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition, 40(7), 2038-2048. https://doi.org/10.1016/j.patcog.2006.12.01



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY