Human Stress Detection Based on Sleeping Habits Using Machine Learning Algorithms

J. G. Jayawickrama
Department of Economics and Statistics
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
gayarajayawickrama96@gmail.com

R.A.H.M. Rupasingha
Department of Economics and Statistics
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
hmrupasingha@gmail.com

2023 3rd International Conference on Advanced Research in Computing (ICARC) | 979-8-3503-4737-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICARC57651.2023.10145663

Abstract— **Stress is a mental or emotional state brought on by demanding or unavoidable circumstances, also referred to as stressors. In order to prevent any unfavorable occurrences in life, it is crucial to understand human stress levels. Sleep disturbances are related to a number of physical, mental, and social problems. This study's main objective is to investigate how human stress might be detected using machine learning algorithms based on sleep-related behaviors. The obtained dataset includes various sleep habits and stress levels. Six machine learning techniques, including Multilayer Perception (MLP), Random Forest, Support Vector Machine (SVM), Decision Trees, Naïve Bayes and Logistic Regression were utilized in the classification level after the data had been preprocessed in order to compare and obtain the most accurate results. Based on the experiment results, it can be concluded that the Naïve Bayes algorithm, when used to classify the data, can do so with 91.27% accuracy, high precision, recall, and f-measure values, as well as the lowest mean absolute error (MAE) and root mean squared error rates (RMSE). We can estimate human stress levels using the study's findings, and we can address pertinent problems as soon as possible.**

Keywords—Classification, Machine Learning, Sleeping Habits, Stress Detection

I. INTRODUCTION

A specific strain exerted on the human body as a result of a number of stimuli is also known as stress. The human body releases stress hormones when it is under stress. Absolute, physiological, relative stressors, and psychological are all different types of stressors [1]. Additionally, stress not only affects your attitude, your relationships, energy level, and duty performance, but it can also cause or aggravate a variety of medical conditions. Therefore, sleep is an essential component in preserving human homeostasis. Sleep disturbances are related to a number of physical, mental, and social problems. In many nations, chronic sleep deprivation is becoming a common. Because the body's stress systems are so essential in adjusting to a continuously moving and stressful background, it's crucial to understand how sleep loss affects them. The human body activates its defenses systems in an adaptive effort to maintain homeostasis. Insomnia may happen if these protections are ineffective. A disruption in routine, such as a psychiatric condition, an illness, or worry, can result in short-term insomnia [2].

The major goal of this study is to detect how human stress change based on sleeping habits. Also the specific objectives are identify the benefits of using a model to detect human stress, identify the relationship between human stress and sleeping habits, identify the major human sleeping behaviors that affect a person's stress, identify the techniques we can use for detecting stress of human and finally, detecting the stress of human based on sleeping habits.

We used data mining methods to build a model to determine the level of stress of the people before and after sleep. It is a tool that enables the study, examination, and visualization of extremely large data sets at a high level of abstraction in data mining. Six different data mining techniques, including Decision Trees, Naïve Bayes, SVM, Random Forest, MLP, and Logistic Regression, are utilized to categorize stress levels. Data can be analyzed by these ML algorithms to predict whether or not people would feel worried the next day.

Here, we can assess human's sleep habits and degrees of stress while teaching them how to find human stress before, after and during sleep. The bulk of available studies predicts human stress in and through sleep using just a few independent variables. However, it is difficult to forecast human stress when there are so few independent variables. By extending the number of attributes in our study, we were able to consider many more. respiration rate, Snoring range, limb movement rate, body temperature, eye movement, blood oxygen level, sleep time, heart rate, and stress levels are just few of the factors to consider. Also we increased the number of classification models to overcome the limitations of existing approaches. This study looks into several taxonomic techniques for predicting human stress in and during. And the main objective of this study is to detect the way of human stress change based on sleeping habits.

In this study, successfully outperform Decision Tree (J48), Random Forest, SVM, MLP and Logistic Regression results by employing a Naive Bayes machine-learning algorithm. We collect information on human behaviors based on the categories listed above. The optimal method among these six algorithms for the suggested strategy was identified based on accuracy, precision, f-measure, recall, RMSE and MAE values. Then take the 10-folds cross validation with the maximum accuracy by using the WEKA data mining tool. The paper's reminding is structured as follows. We go over similar work in section II. The methodology of the research is discussed in Section III. The Section IV is discussed evaluation and experiments, and Section V discusses the future work.

II. LITERATURE REVIEW

A. Related Work

Stress is regarded as a mental cancer [3], and it has become a public health crisis in the twenty-first century [4]. It is also believed to be one of the unique qualities of our lives, and everyone encounters it at some point in their lives as an unavoidable aspect [5]. Sleep is essential for both physiological and psychological well-being. Sleep deprivation has been linked to an increased risk of cardiovascular illness, type 2 diabetes, cognitive issues, attention deficit hyperactivity disorder (ADHD), depression, and poor performance. Based on the study [6] poor sleep has a greater impact on the mood of the next day. Most importantly, here we are predicting how human stress detects through sleep.

The study [7] has presented a method to detect and admin physiological stress about food habits on the Internet of Medical Things (IoMT). An addition to these works, they offer SaYoPillow, which monitors and controls a person's stress levels as they sleep. SaYoPillow's main objective is to achieve "Smart-Sleeping," which is a comprehensive sleep that satisfies the optimal bodily needs for sleep. SayoPillow suggested a real-time physiological signal detecting to admin sleep quality by taking into account parameters such as: respiratory rate range, the number of hours of sleep, heart rate range, snoring range, eye movement rate, oxygen in blood range, or duration of time spent in Rapid Eye Movement (REM), change in body temperature and limb movement rate. The likelihood of experiencing tension and other health issues rises when snoring rates are above 50dB. A good breathing rate is regarded as being between 15 and 17 breaths per minute (bpm). A person's heart beats 5–10 times more slowly while they are sleeping than normal. Adults should get at least 7 hours of sleep per night because sleep deprivation is harmful to their health. Then it is recommended that 20-25 percent of total sleep duration be spent in (REM), which is around 90 minutes for 7-8 hours of sleep. Low oxygen levels are considered aberrant and stressful when they fall below 90%.

Based the study [8], the majority of car accidents are caused by stress. When a driver is in this emotional state, he or she makes more errors on the road. They suggested a method for determining the appropriate speed for each road segment in this research based on the degree of stress. The solution provides feedback following the completion of a road segment. The best average performance is found using MLP and particle swarm optimization (PSO). On Android smartphones, the solution was implemented. According to the findings, when drivers apply the notion, they drive more smoothly and with less stress. The input variables include average speed, standard deviation of vehicle speed, average acceleration, positive kinetic energy, weather, traffic, time, and weekday.

In the study ML framework for the monitoring mental stress at multiple levels [9], 348 individuals between the ages of 20 and 60 were involved, both men and women, employed and unemployed, and carrying out a range of tasks from professional duties to domestic chores. Based on their result, Random Forest had the highest f1 score with 90% for stress and Naive Bayes for depression.

The study [10] proposed an ML framework based on the electroencephalogram (EEG) signal analysis of stressed peoples. A popular assessment method based on the Montreal Imaging Stress Task was used to induce stress in the laboratory. The introduction of stress was supported by both the work performance and the objective feedback. The suggested ML model included EEG feature extraction and selection (using the t-test, ROC curve, and Bhattacharya distance), classification with SVM, logistic regression, regression analysis and Naive Bayes models. According to the findings, the proposed framework generated 94.6% accuracy for level-two stress detection and 83.4% accuracy for multiple-level detection.

The review study [11] was completed in a number of stages, which included data gathering using the Web of Science database's closest keywords and network visualization design based on prior data. To analyze the work, they used the four closest keywords, five study papers, three publishers, and four journals. The outcomes demonstrated that SVM successfully classified the signals and had a system sensitivity of 91.18%. This study provided a more thorough and significant description of the path that future research will take.

According to studies [12], the most prevalent physiological marker for determining stress is the galvanic skin response (GSR). GSR has been linked to arousal, both psychological and physiological. As a outcome of greater sweat gland activity brought on by ANS arousal, skin conductance increases. They emphasized the importance of constant, on-going stress tracking in this circumstance. They investigated the methodologies for study, tracking, and stress prediction. They also discussed earlier research that tracked real-time stress using an edge computing framework.

The research [13] found that as the prevalence of wearable, sensor-enabled portable, and implantable devices increases in the developing Internet of Things (IoT), physiological sensor analytics is developing into a crucial tool for health monitoring. The identification of stress has previously been accomplished through physiological multisensor studies. In order to create a reliable and effective system for precise stress identification, they concentrated on ECG monitoring, which is now possible using slightly invasive wearable patches and sensors. They take into account eight different kinds of input data: electromyogram (EMG), ECG hand marker, GSR time stamp, foot galvanic skin response (GSR), intermittent heart rate (IHR), and breathing. In this study, researchers use machine learning techniques and algorithms to identify signs of stress in automobile drivers who are experiencing varying degrees of external stress brought on by road conditions.

The study [14] describes a method for evaluating pilgrims' stress levels by analyzing their nighttime sleep patterns and finding the most significant sleep metrics for stress identification. They develop and evaluate different classification models using bio-physiological indicators like ECG/HRV, respiration, body temperature, and GSR data, as well as physical traits like upper body position sensors and accelerometers on the body and arms. They were able to create person-independent models that distinguished between three stress levels: low, moderate, and high using the classification models. They achieved the best classification accuracy of 73 % using SVM than other algorithms.

Describe the summary of related works with methodology, results, and their limitations in the following TABLE I.

B. Gaps in Literature

Existing studies on health and stress did not specifically address how stress impacted people's sleeping patterns.

TABLE I RELATED WORKS SUMMERY

Ref. No	Major attributes	Classifier/ Method	Findings	Limitations
[7]	Sleeping habits	IoT cloud	A cutting-edge device called the Smart-Yoga Pillow (SaYoPillow) is being suggested in order to fully realize the idea of "Smart-Sleeping" and to help with understanding the connection between stress and sleep.	Used IoT cloud method only and the EHR systems are currently centralized and vulnerable to a single point of failure.
[8]	Average speed and acceleration, Standard deviation of vehicle speed, Positive kinetic energy, Weather and traffic state, Time, Weekday	Leaner Regression, SVM, and Naïve base	Proposed an algorithm to estimate the optimum speed based on the stress level for each road section.	Through visual inspection, the EEG signal was manually cleaned of artifacts.
[9]	User contextual information	Random forest, SVM, K-Nearest Neighbor (KNN)	An ML system that analyzes participants' stressed-out EEG signals is proposed.	Presents a portion of the relevant literature for several of the individual cases but does not provide an exhaustive analysis of each of the individual cases.
[13]	Eight features using ECG signals	SVM, Linear discriminant analysis and Naïve base, KNN	Examine and discuss the studies that have employed machine learning-based stress detection algorithms.	Focused on psychological stress. However there is still a need for more research into physiological stress detection.
[14]	RESP, ECG/HRV, GSR data, body temperature and and physical parameters	Leaner Regression, SVM, Random forest, Neural Network, KNN	The most relevant sleep measures were found using a variety of physical and bio-physiological features from wristband sensors and chest strap devices to differentiate between high, low, and moderate stress.	They could only identify changes in sleep postures involving the upper body using the Zephyr posture sensor and there are just ten subjects available.

Additionally, the current research does not make any predictions; it just examines the connection between stress and sleeping patterns. We couldn't find any research using ML systems to analyze sleep patterns before, after and during sleep to determine human stress. The most of obtainable research use only a few independent factors to predict human stress in and through sleep. However, with only a few independent factors, detecting human stress is challenging. By extending the number of attributes in our study, we were able to increase the accuracy.

III. RESEARCH METHODOLOGY

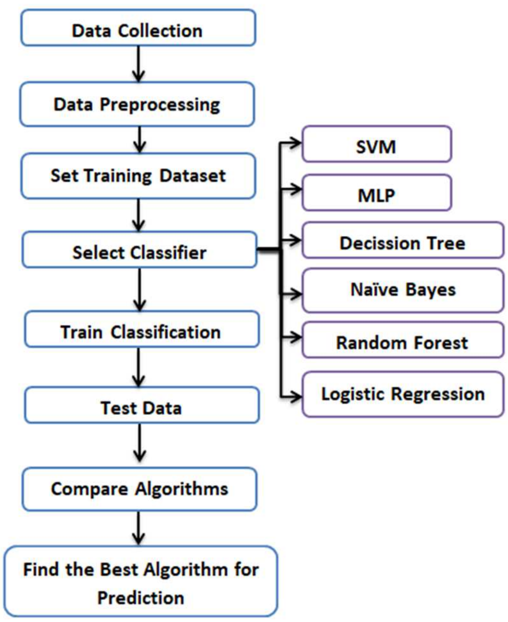

Fig. 1. Architecture of the study

This research mainly focuses to predict human stress based on the behaviors in sleep. The proposed approach includes main three steps. They are data collection, data preprocessing, and classification. Following Fig. 1 explains the architecture of the study.

A. Data Collection

After the study problem has been identified we used secondary data for the research approach. The dataset was obtained from the 'Kaggle' [15] website and was obtained under the seven sleeping habits. The scope of the study was defined as people in society and around 500 people responded to the study. According to the dataset limb movement, snoring range, body temperature, respiration rate, eye movement, blood oxygen level, and heart rate are considered as seven factors. It shows in Fig. 2. These factors aid in determining the association or her stress level

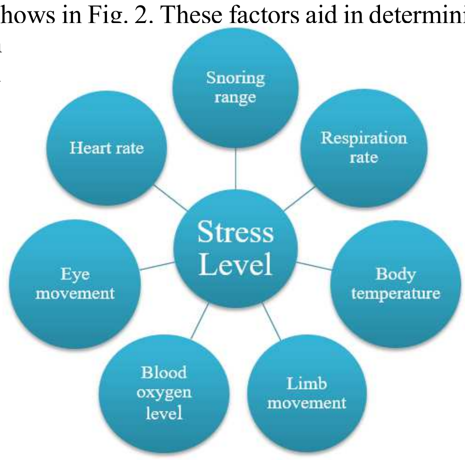

Fig. 2. Attributes in the sample

B. Data preprocessing

In this procedure, incomplete and irrelevant data is identified, and it is then replaced, amended, or eliminated unless it is replaced, altered, or removed. The raw dataset is cleared using the data preparation technique to remove inaccurate, unreliable, and poor data. The data is pre-processed using data cleansing, data transformation, and data reduction. The WEKA software uses a ranking of the attributes to determine which are most impacted by the outcome. According to the ranker among all independent variables, the snoring range has more influence on the dependent variable. The respiration rate has the next greatest effect on the dependent variable. Also sleeping hours have the lowest effect on the dependent variable. Finally snoring range, body temperature, respiration rate, heart rate, blood oxygen level, limb movement and eye movement are used as seven independent variables. The stress level is used as a dependent variable. The collected original dataset contained four different stress levels but we combined the first and second stress levels as the low-level stress and the third and fourth stress levels as the high-stress level for our approach. Also, most of the decimal numbers in the dataset are rounded to the nearest decimal point or whole numbers.

C. Classification

The WEKA software is used to classify and assess the pre-processed data. The input data collection is kept in CSV file format. The data collection is divided into training-testing data in 80% and 20% groups. The predictions are created using a variety of algorithms from the WEKA softwareThe model was created to target human stress based on sleep behaviors that individuals described as 1- low/normal, 2- high. The decision tree, Naive Bayes, SVM, Random Forest, logistic regression, and MLP classification algorithms were used to identify the best classification model using the testing data after the training and testing data had been divided.

- Decision tree: A tree structure is used to model the many links between the features and the potential output data in decision trees, which are effective algorithms for classifying data.
- Random Forest: It is a classifier that employs numerous decision trees on different input dataset groups and took averages the outcomes to increase the predicted accuracy of the given dataset.
- SVM: Its goal is to determine the best line or decision boundary that can categorize n-dimensional space, enabling us to rapidly grouped new data in the future.Naïve Bayes: The value that produces the highest probability after being computed from a conditional probability chain is determined using Naïve Bayes.
- Logistic regression: When you need to forecast the existence or absence of a characteristic or outcome based on the values of a group of predictor variables, this might be helpful.
- MLP: The directed graph that joins the output and input layers of an MLP has many tiers of input nodes.

The two classification groups used are Low and High stress levels. Naïve base showed the best accuracy based on the results. It is a probabilistic ML model that is used to solve classification problems. The categorization is based on the Bayesian theorem and the assumption that predictors are independent. Fig. 3 explain the process of getting results.

IV. RESULTS & DESCUSSIONS

Microsoft Windows 10 was installed on a PC with an Intel® Core (TM) i5-5005U CPU running at 2.00GHz and RAM of 4.0 GB. The testing/training environment is the WEKA 3.9.5 software. The collected 504 data using a Kaggle website [15] was used for a data set for the evaluation and experiment process.

In this work, Decision Tree (J48), Nave Bayes, MLP, Random Forest, SVM, and Logistic Regression were used for comparison and evaluation. The accuracy outcomes of these algorithms employing cross validation are shown in Fig. 4.

The WEKA software employs cross-validation (by 10) to manage test results and determine the precision of each strategy. The Naïve Bayes classifier has the better accuracy when checking with other algorithms. The Naïve Bayes model, which has an accuracy of 91.27%, is the best one for forecasting human stress.

The recall, precision, and f-measure computations were then used to compare the evaluation findings, as shown in (1), (2), and (3). These are utilized to check the suitability of the results. The values for six algorithms are shown in TABLE II.

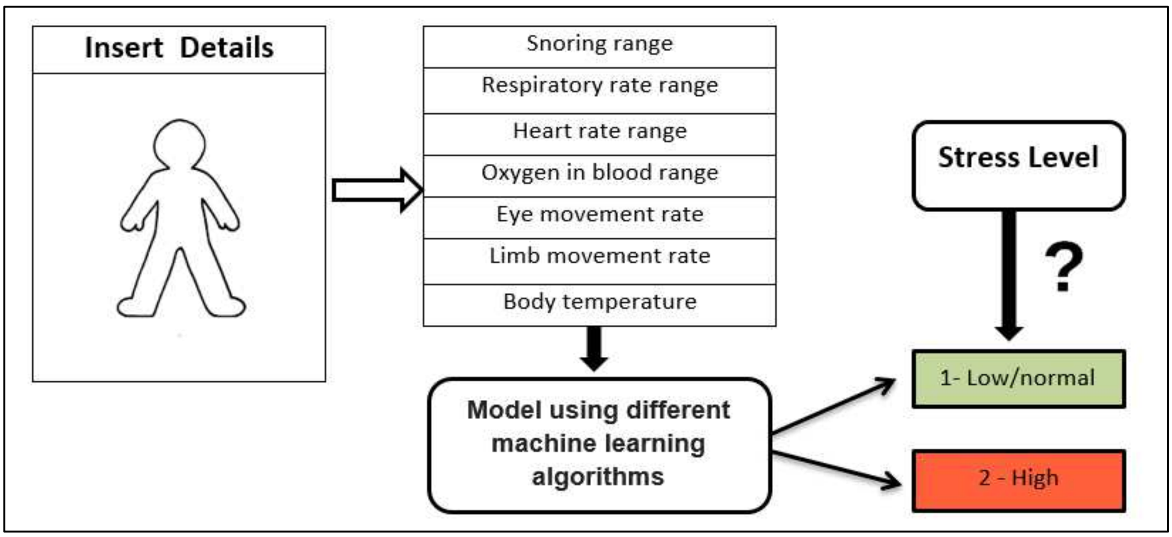

Fig. 3. Process of getting results

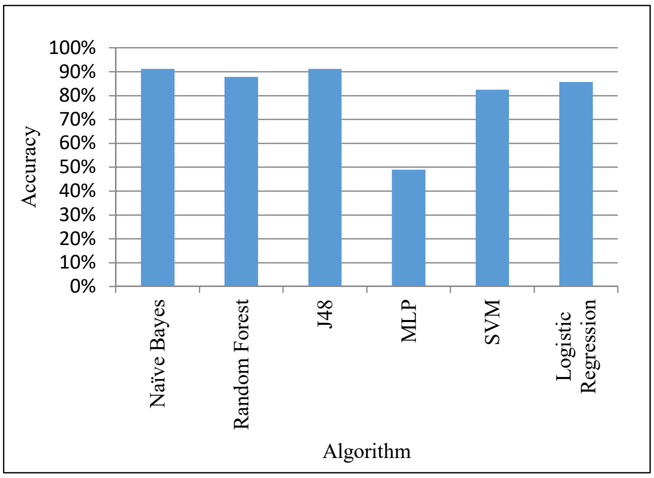

Fig. 4. Aaccuracy results of these algorithms

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + Fasle\ Negative} \quad (2)$$

$$F - Measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3)$$

The findings show that when compared to all other classification methods, the Naïve Bayes algorithm has the highest recall, precision, and f-measure values.

TABLE II RECALL, PRECISION AND F-MEASURE OF ALGORITHMS

Algorithm	Precision	Recall	F-measure
Naïve Bayes	0.916	0.916	0.916
Random Forest	0.879	0.879	0.879
Decision Tree	0.916	0.913	0.913
MLP	0.501	0.49	0.486
SVM	0.847	0.825	0.82
Logistic Regression	0.857	0.857	0.857

The lowest error, as determined by the MAE and RMSE, increases the algorithm's accuracy. Following (4) and (5), respectively, describe these two requirements. The sample size is n, the real value is P_I, and the anticipated value is Q_I. The lowest RMSE is found in the decision tree algorithm and the lowest MAE is found in the Naïve Bayes algorithm.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | P_i - Q_i | \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (P_i - Q_i)^2} \quad (5)$$

Fig. 5 explain the RMSE and MAE values for the six algorithms.

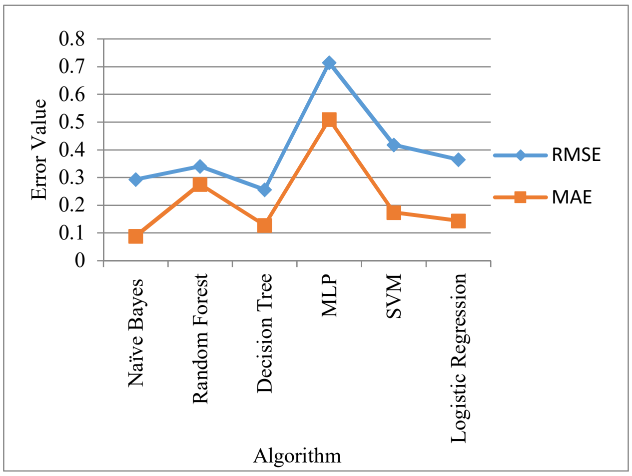

Fig.5. RMSE and MAE error values

The evaluations made above indicate that Naive Bayes is the best algorithm for predicting human stress. It has the best RMSE and MAE error values, as well as the best accuracy, precision, recall, and f-measure values. Additionally, the Decision Tree has lower error rates and greater accuracy. But in this research, Naive Bayes has demonstrated outstanding accuracy when compared to all other methodologies.

The comparison between 5-fold and 10-fold cross validation are shown in TABLE III. The results of the both cross-validations were very similar.

TABLE III. RESULT OF 5- VS. 10- CROSS VALIDATION FOLDS

Folds	Precision	Recall	F Measure	MAE	RMSE
5	0.916	0.913	0.913	0.0885	0.2942
10	0.916	0.916	0.916	0.0884	0.2938

TABLE IV discusses the confusion matrix for above six algorithms.

TABLE IV CONFUSION MATRIX

Category	TP	TN	FP	FN
Naïve Bayes	240	220	12	32
Random Forest	240	200	32	32
Decision Tree	240	220	12	32
MLP	109	138	94	163
SVM	272	76	156	0
Logistic Regression	237	198	34	35

The accuracy of the models is tested based on sleeping habits, as compared to existing approaches. In comparison to other methodologies, this study reveals that Naïve Bayes has a high level of accuracy.

V. CONCLUSION

Human stress is depending on the different criteria and it is important to understand the human stress level to avoid some unnecessary problems. The purpose of this study is to detect how human stress change based on sleeping habits. Also, we identified the benefits of using a model to detect human stress and the connection between human stress and sleep. For this purpose, we collected data including human stress levels and seven habits as the variables through sleep. We used six alternative machine learning algorithms such as Random Forest, MLP, Logistic Regression, Decision Tree, Naïve Bayes, and SVM were evaluated using sleeping habits and human stress level. The evaluation is done 10-fold cross-validation. Based on the evaluation results, Naive Bayes outperforms the other five algorithms, and it is the most effective in forecasting human stress. The best recall, precision, and f-measure values, as well as a lower error rate in MAE and RMSE values, go hand in hand with the Naive Bayes method's 91.27% accuracy. When discussing the usability and potential applications of this work, we may utilize this model to detect the stress level of the people by adding their sleeping habits to the model, which served as the study's independent variables. Based on the result of stress level, we can deal with that person.

In the future, to increase the accuracy of the results, we intend to multiply the data and employ the ensemble learning method, which combines all six algorithms. Because of the less data we could not use the neural networks and deep learning techniques here. Therefore, we intend to apply those algorithms by expanding our dataset, and then it uses to boost the existing accuracy.

REFERENCES

[1] N. Schneiderman, G. Ironson, and S. D. Siegel, "Stress and health: Psychological, behavioral, and biological determinants," Annu. Rev. Clin. Psychol., vol. 1, pp. 607–628, 2005, doi: 10.1146/annurev.clinpsy.1.102803.144141.

[2] 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.

[3] 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.

[4] G. Fink, "Stress : Concepts , Cognition , Emotion , and Behavior : Handbook of Stress Stress : Concepts , Definition , and History George Fink Florey Institute of Neuroscience and Mental Health," no. October, 2017, [Online]. Available: https://www.researchgate.net/profile/George-Fink/publication/317026245_Stress_Concepts_Cognition_Emotion_and_Behavior_Handbook_of_Stress/links/59d17f1b0f7e9b4fd7fa28b3/Stress-Concepts-Cognition-Emotion-and-Behavior-Handbook-of-Stress.pdf

[5] A. M. Shahsavarani, E. A. M. Abadi, and M. H. Kalkhoran, "Stress Assessment and Development of a Primary Care of Psychology Service," Int. J. Med. Rev., vol. 2, no. 2, pp. 230–241, 2014.

[6] A. Khademi, Y. El-Manzalawy, L. Master, O. M. Buxton, and V. G. Honavar, "Personalized sleep parameters estimation from actigraphy: A machine learning approach," Nat. Sci. Sleep, vol. 11, pp. 387–399, 2019, doi: 10.2147/NSS.S220716.

[7] L. Rachakonda, A. K. Bapatla, S. P. Mohanty, and E. Kougianos, "SaYoPillow: Blockchain-Integrated Privacy-Assured IoMT Framework for Stress Management Considering Sleeping Habits," IEEE Trans. Consum. Electron., vol. 67, no. 1, pp. 20–29, 2021, doi: 10.1109/TCE.2020.3043683.

[8] V. C. Magana and M. Munoz-Organero, "Reducing stress on habitual journeys," 5th IEEE Int. Conf. Consum. Electron. - Berlin, ICCE-Berlin 2015, pp. 153–157, 2016, doi: 10.1109/ICCE-Berlin.2015.7391220.

[9] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, "Machine learning framework for the detection of mental stress at multiple levels," IEEE Access, vol. 5, no. c, pp. 13545–13556, 2017, doi: 10.1109/ACCESS.2017.2723622.

[10] E. Garcia-Ceja, M. Riegler, T. Nordgreen, P. Jakobsen, K. J. Oedegaard, and J. Tørresen, "Mental health monitoring with multimodal sensing and machine learning: A survey," Pervasive Mob. Comput., vol. 51, pp. 1–26, 2018, doi: 10.1016/j.pmcj.2018.09.003.

[11] F. Akhtar, M. B. Bin Heyat, J. P. Li, P. K. Patel, Rishipal, and B. Guragai, "Role of Machine Learning in Human Stress: A Review," 2020 17th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2020, pp. 170–174, 2020, doi: 10.1109/ICCWAMTIP51612.2020.9317396.

[12] 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.

[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.

[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.

[15] Kaggle, https://www.kaggle.com/ (accessed: Feb. 27, 2022).