INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

Stress Detection Using Machine Learning Algorithms

Dr Usha Kamale

Professor, ECE Department, MVSR Engineering College, Hyderabad, Telangana, India

DOI: https://doi.org/10.51244/IJRSI.2026.1304000062

Received: 08 April 2026; Accepted: 13 April 2026; Published: 29 April 2026

ABSTRACT

Stress management is becoming more and more crucial in today's fast-paced technological environment, particularly for IT professionals. Long working hours, strict deadlines and high expectations are common aspects of the work environment in the IT sector, and these can raise stress levels. Unmanaged stress has an adverse effect on professionals' health and well-being as well as their productivity and job happiness. A data set comprising of 2343 sample values taken from Kaggle is used for detecting the stress levels.

Keywords— Stress detection, Machine learning, Deep Neural Networks, Random forest, Ada-boost, Neural networks, extra trees

INTRODUCTION

This paper presents the prediction of the stress levels of IT professionals using machine learning techniques, leveraging features such as Heart Rate, Skin Conductivity, Hours Worked, Number of Emails Sent, and Meetings Attended. These features encompass both physiological and work- related factors contributing to stress, offering a comprehensive view. By applying machine learning, the study innovatively addresses the pressing issue of work place stress, utilizing data analytics to generate actionable insights for individuals and organizations. Individuals can use these predictionsfor self-monitoring and early intervention, while organizations can identify high-stress environments or roles to allocate resources or interventions more effectively.

Preliminary results show a strong correlation between the selected features and stress levels, validating the use of machine learning for stress prediction in IT professionals. This approach demonstrates the model's potential to enhance mental health and well-being in the workplace by enabling timely and targeted interventions. By providing a data-driven method for stress management, the study contributes significantly to improving productivity and job satisfaction in the IT industry.

Objectives

- To detect stress using parameters such as Heart rate, Skin conductivity, Hours worked, Emails sent, Meetings attended.
- To predict the stress levels based on data set from the Kaggle using Random forest regressor, Ada boost, Extra Tree algorithms, and Neural network.

METHODOLOGY AND APPROACH

Existing Method

In existing methods, K-Means clustering has been employed to segment IT professionals into different stress categories based on similar behavioral and physiological traits. Principal Component Analysis (PCA) is often used as a dimensionality reduction technique to simplify the dataset, making it easier to apply subsequent machine learning algorithms like Logistic Regression. Based on a mix of physiological and work-related characteristics, individuals can be divided into stressed and non-stressed groups using the commonly used predictive model known as logistic regression.

K-means clustering: K-means clustering is an unsupervised machine learning algorithm that partitions a dataset into 'k' distinct, non-overlapping subsets (clusters) based on similarity. The algorithm aims to minimize the variance within each cluster and maximize the distance between clusters. The data set is divided into 'k' clusters such that each data point is closer to the centroid of its assigned cluster than to any other cluster centroid. Advantages of K-means clustering are Fast and efficient, it also provides strong coupling between data points within a cluster, and it is suitable for numerical data where distance is meaningful. The limitations are it is Sensitive to initial cluster assignments, it may get stuck in local minima and not suitable for datasets with overlapping clusters.

K-means clustering is widely used in various applications, including customer segmentation, image segmentation, and feature learning in machine learning models. The performance of K-means clustering for stress detection can be evaluated using metrics such as accuracy, precision, recall, and F1-score. Studies have reported promising results using K-means clustering for stress detection, with high accuracy and precision achieved in some cases.

In this study, an electro cardiogram signal is used to check the state of stress. Feature points according to the difference between the R-R interval and the R-S peak value are extracted from the acquired ECG signal. Finally, K- means clustering is applied and classified into cases of stress and no stress.

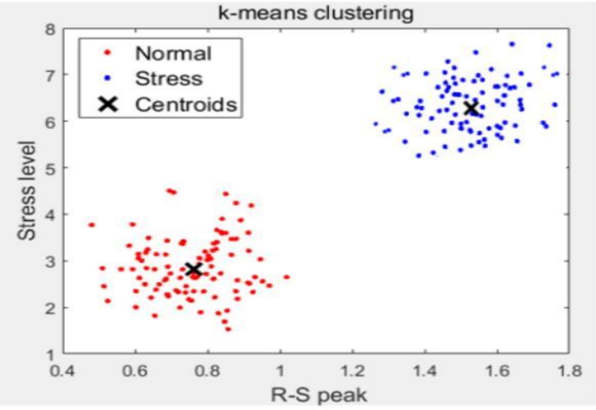


Fig. 1 Stress Intensity as per R-S peak value

The above figure shows the stress intensity which is expressed according to the R-S peakvalue.

The main challenges of using K-means clustering for stress detection include:

Initial Placement of Centroids: The K-means algorithm is sensitive to the initial placement of centroids, which can significantly affect the quality of the clustering results. This can be particularly challenging when dealing with physiological signals that are inherently noisy and may not be well-separated.

Number of Clusters: Choosing the optimal number of clusters (K)is crucial for effective stress detection. However, this can be difficult, especially when dealing with complex physiological signals that may exhibit multiple patterns related to stress.

Noise and Outliers: Physiological signals can be noisy and contain outliers, which can negatively impact the performance of the K-means algorithm. Effective methods for handling noise and outliers are essential for reliable stress detection.

Interpretability: K-means clustering can be challenging to interpret, especially when dealing with complex physiological signals. It is essential to develop methods for effectively communicating the results of the clustering algorithm to users.

Data Quality: The quality of the physiological signals used for stress detection is critical. Poor data quality can

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

lead to inaccurate clustering and ineffective stress detection.

Scalability: As the amount of data increases, the computational complexity of the K-means algorithm can become a significant challenge. Scalable methods for handling large datasets are essential for practical applications.

Handling High-Dimensional Data: Physiological signals often involve high-dimensional data, which can be challenging to handle using traditional K-means clustering. Techniques for handling high-dimensional data, such as dimensionality reduction, are essential for effective stress detection.

Labeling and Annotation: Manually labeling and annotating physiological signals as stress or non-stress can be time-consuming and prone to inaccuracies. Developing methods for automated labeling and annotation is essential for practical applications.

These challenges high-light the need for careful consideration of the algorithmic and methodological aspects of K-means clustering for stress detection, as well as the importance of addressing these challenges through innovative solutions.

Proposed Method

In the proposed system, we leverage ensemble machine learning techniques like Random Forest, Ada-Boost, and Extra Trees to predict stress levels in IT professionals. These sophisticated models provide a more detailed understanding of stress variables by capturing the complex interactions between many physiological and work-related aspects. By employing ensemble methods, the system aims to achieve higher predictive accuracy and robustness compared to traditional methods like K-Means and Logistic Regression.

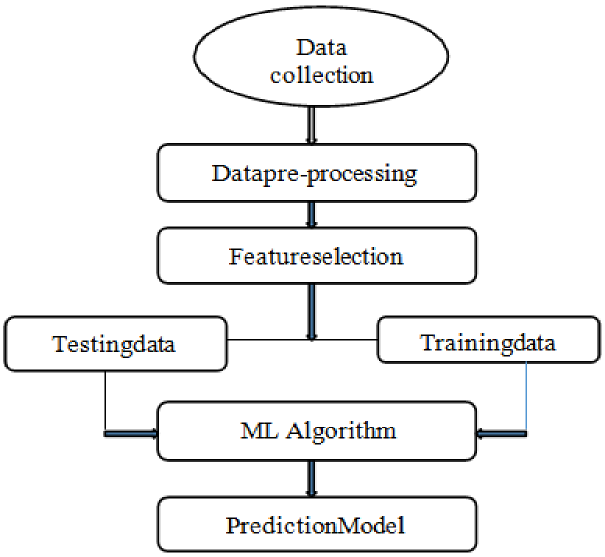


Fig. 2 Stress Detection algorithm

The proposed method of stress detection using the Random Forest algorithm involves the following steps:

Data Collection: Physiological signals such as heart rate, RR interval, and Galvanic skin response are collected from participants using wearable devices or sensors.

Data pre-processing: It involves transforming raw data into a format that is consumable, understandable, and usable for further analysis. The primary goal of data pre-processing is to improve the quality of the data and make it more suitable for the specific data mining task. The key steps involved in data preprocessing:

Data Cleaning: This step involves identifying and correcting errors or inconsistencies in the data, such as missing values, outliers, and duplicates. Techniques used for data cleaning include imputation, removal, and transformation.

Data Integration: This step involves combining data from multiple sources to create a unified dataset. Because it involves handling data with various formats, structures, and semantics, data integration can be difficult. Techniques such as record linkage and data fusion are used for data integration.

Data Transformation: This step involves converting the data into a suitable format for analysis. Normalization, standardization, and discretization are common methods used in data transformation. Normalization is used to scale the data to a common range, while standardization is used to transform the data to have zero mean and unit variance. Continuous data can be discretized using the discretization process.

Data Reduction: This step involves reducing the size of the dataset while preserving the important information. Techniques like feature selection and feature extraction can be used to reduce data. While feature extraction entails converting the data into a lower-dimensional space while maintaining thecrucialinformation, feature selection entails choosing a subset of pertinent characteristics from the dataset.

Data Discretization: This step involves dividing continuous data into discrete categories or intervals. When developing algorithms for data mining and machine learning that need categorical data, discretization is frequently utilized. Techniques such as equal width binning, equal frequency binning, and clustering are used for discretization.

Data Normalization: This step involves scaling the data to a common range, such as between 0 and 1 or -1 and 1. In order to manage data with various scales and units, normalization is frequently utilized. Decimal scaling, z-score normalization, and min-max normalization are examples of common normalization methods.

In order to guarantee the accuracy of the analysis results and the quality of the data, data preparation is essential. The precise processesin datapreprocessing can changebasedon the type of data and the objectives of the study. By performing these steps, the data mining process becomes more efficient and the results become more accurate.

Feature selection: It is a crucial step in machine learning that helps to improve the performance and accuracy of the model by selecting the most relevant features. By understanding the different types of feature selection techniques and popular methods, you can effectively apply feature selection to your machine learning projects and achieve better results.

Training Data: Training data is a subset of the original dataset used to train a machine learning model. It is the data thatthemodellearnsfromtomakepredictions.TheTraining data is typically larger than testing data, as it is necessary to feed the model with a sufficient amount of data to learn meaningful patterns

Testing data: Testing data isa subset of theoriginal dataset used to evaluate the performance of a trained machine learning model. It is the data that the model is tested against to ensure it generalizes well to new, unseen data. The Testing data is typically smaller than training data, as it is used to evaluate the model's performance on a smaller, representative subset of the data.

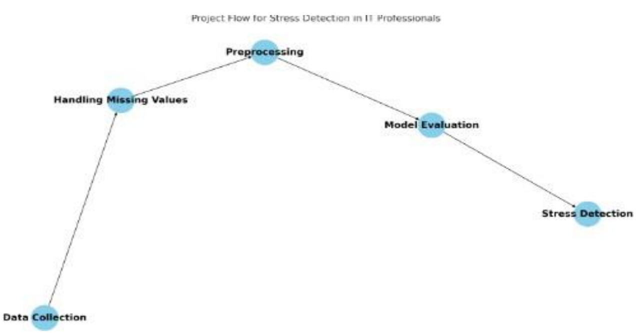


Fig. 3 Stress Detection work flow diagram

Random Forest Algorithm: The extracted features are used as input to the Random Forest algorithm, which is a type of ensemble learning method that combines multiple decision trees to make predictions.

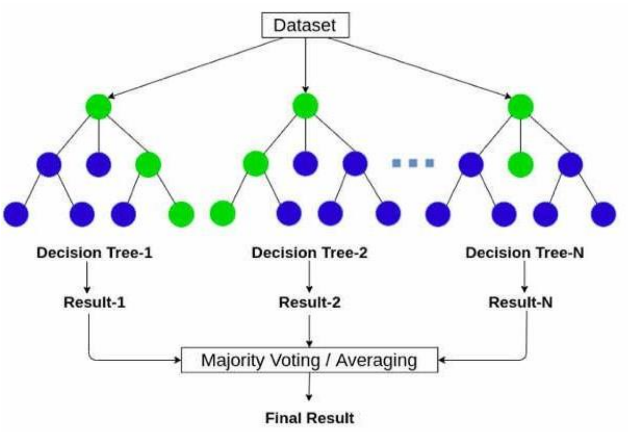
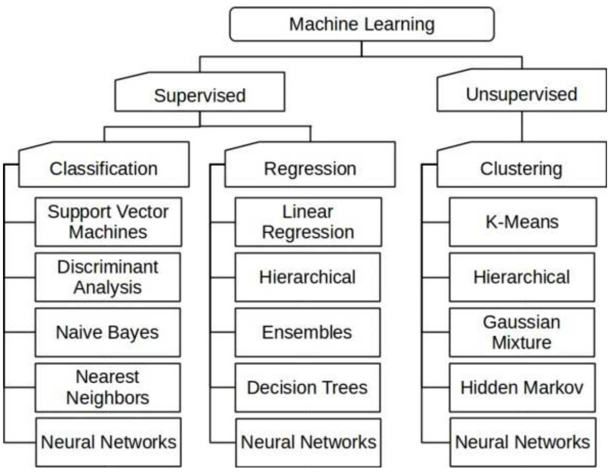


Fig. 4 Random Forest Stress Detection work flow diagram

The Random Forest algorithm is chosen for this task because of its ability to handle complex data interactions and provide robust predictions. The algorithm is also effective in reducing over fitting and improving the overall accuracy of the model. It is well-suited for stress detection due to its ability to handle complex datasets and provide reliable predictions. The Random Forest Regressor can achieve very high accuracy and compared to other algorithms like logistic regression and decision trees, random forest has been found to out-perform in terms of accuracy for stress detection tasks.

Machine Learning

There are three primary types of machine learning algorithms:



Supervised Learning: Supervised machine learning is a type of machine learning where the algorithm is trained on labeled data to make predictions or decisions based on the data inputs. The algorithm learns from the labeled data to identify patterns and relationships between the input features and the target output. We call this fitting or training process. This type of learning involves training models on labeled data sets to learn a mapping from input data to output labels.

Labeled Data: Supervised learning uses labeled data, where each input data point is associated with a correct output or target value. This labeled data is used to train the algorithm to learn the relationship between the input features and the target output.

Supervised learning can be divided into two main types of problems:

Regression: Predicts continuous target variables, such as linear regression, logistic regression, and support vector machines (SVMs).

Classification: Predicts categorical target variables, such as logistic regression, support vector machines(SVMs), random forest, decision tree, and k-nearest neighbors (KNN).

Common supervised machine learning algorithms include:

Linear Regression: Predicting continuous values based on a linear relationship between features and the target output.

Logistic Regression: Predicting categorical outcomes using a logistic function to calculate the probability of the target output.

Support Vector Machines (SVM): Creating a hyper-plane to segregate n-dimensional space into classes and identify the correct category of new data points.

K-Nearest Neighbors (KNN): Predicting the class or value of a new input based on the majority class or average value of the k nearest neighbors.

The advantages of Supervised learning are:

High Accuracy: Supervised learning can achieve high accuracy in predicting outcomes if the training data is representative and the algorithm is chosen correctly.

Interpretability: Supervised learning models can be interpreted and understood by humans, making them useful for many applications.

Unsupervised Learning: This is a powerful tool in machine learning that can be used to discover hidden patterns and relationships in data without any pre-defined labels or outputs. It has many applications and can be used for a variety of tasks, but it also has some challenges and limitations. It is a type of machine learning where models are trained using unlabeled data and are allowed to act on that data without any predefined labels or outputs. This approach is useful when there is no predefined target variable or when the target variable is not well-defined.

Unsupervised Learning Algorithms:

Clustering: Algorithms for clustering combine related data points into clusters according to their shared characteristics. Dimensionality Reduction: Dimensionality reduction algorithms reduce the number of features in a dataset while preserving the most important information.

Anomaly Detection: Anomaly detection algorithms identify data points that are significantly different from the rest of the data.

Characteristics of Machine Learning Algorithms

Ability to learn from Data

Identifying patterns and making Predictions

Handling various types of data

Improving through experience

Advantages

Enhanced Predictive Accuracy

Data driven Decision Making

Efficiency in Pattern finding

Automation

Real time In-sights

Applications

Machine learning algorithms have numerous applications across various fields. Here are some of the most significant applications:

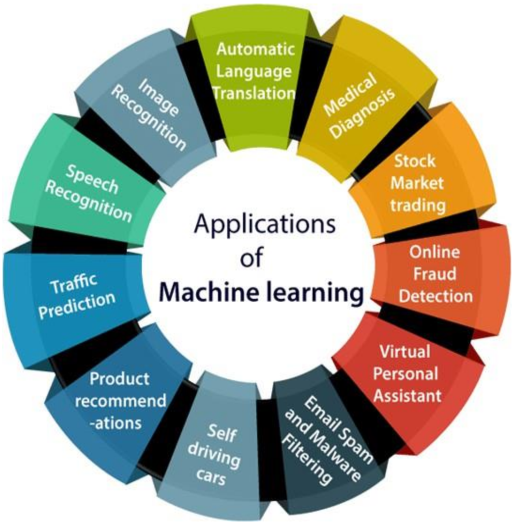


Image Recognition: It is a common application of machine learning. It involves identifying objects, persons, places, and digital images. Examples include Facebook's auto friend tagging suggestion and Google's image recognition features.

Speech Recognition: Speech recognition is another significant application of machine learning. It involves converting voice instructions into text. Examples include Google Assistant, Siri and Alexa.

Traffic Prediction: Traffic prediction is a critical application of machine learning. It involves predicting traffic conditions such as congestion, slow-moving, or heavily congested roads. Google Maps uses machine learning to provide accurate traffic predictions.

Product Recommendations: These are a popular application of machine learning. It involves suggesting products based on customer preferences and behavior. Examples include Netflix's movie and series recommendations and Amazon's product suggestions.

Self-DrivingCars: Self-drivingcarsareasignificant applicationofmachinelearning.Itinvolvestraining cars to detect people and objects while driving. Tesla is working on self-driving cars using unsupervised learning methods.

Email Spam and Malware Filtering: Email spam and malware filtering are critical applications of machine learning. It involves detecting and filtering out spam and malicious e mails. Gmail uses machine learning algorithms to filter out spam emails.

Virtual Personal Assistants:Virtual personal assistants are a popular application of machine learning. It involves using voice instructions to perform tasks. Examples include Google Assistant, Siri, and Alexa.

Online Fraud Detection: Online fraud detection is a critical application of machine learning. It involves detecting fraudulent transactions. Machine learning algorithms are used to identify patterns and detect fraudulent activities.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

Stock Market Trading: Stock market trading is a significant application of machine learning. It involves predicting future price trends and market values. Machine learning algorithms are used to analyze stock market data and make predictions.

Medical Diagnosis: Medical diagnosis is a critical application of machine learning. It involves diagnosing diseases using machine learning algorithms. Examples include breast cancer classification and Parkinson's disease classification.

Automatic Language Translation: Automatic language translation is a popular application of machine learning. It involves translating text from one language to another. Google's Neural Machine Translation (GNMT) is an example of this application.

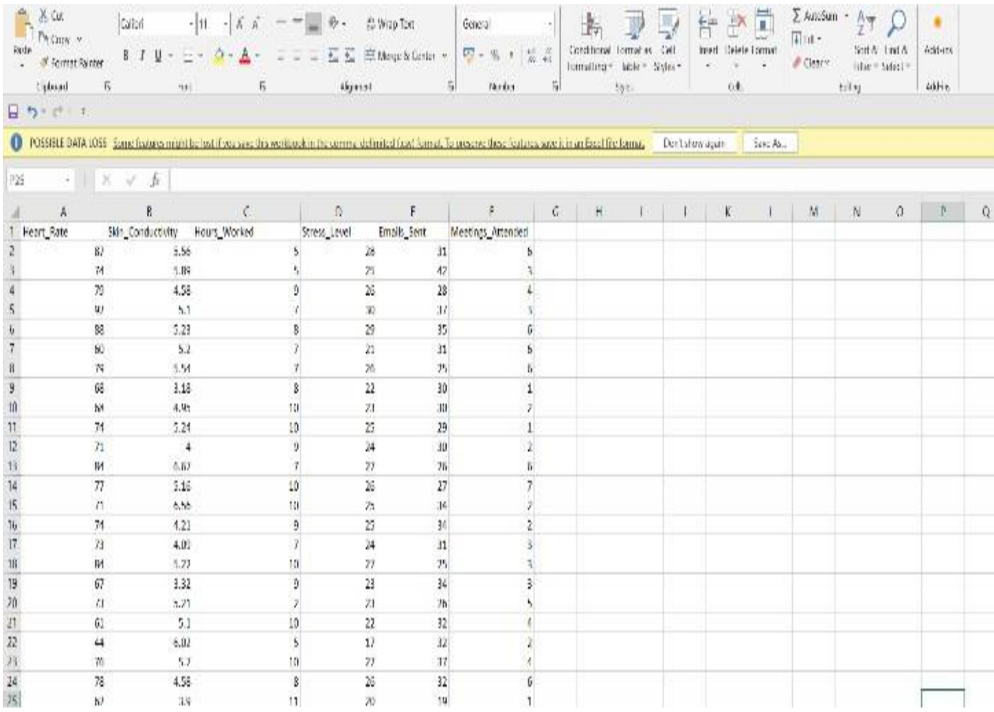
Customer Segmentation: Customer segmentation is an application ofunsupervised learning. It involves grouping customers based on their demographics, behavior, or preferences. This helps businesses understand their customers better and target them with more relevant marketing campaigns.

Fraud Detection: Fraud detection is another application of unsupervised learning. It involves detecting fraudulent transactions by identifying patterns that deviate from the expected norms.

Recommendation Systems: These are the applicationsofunsupervisedlearning.Theyinvolve recommending items to users based on their past behavior or preferences.

Natural Language Processing (NLP): NLP is an application of unsupervised learning. It involves tasks such as topic modeling, document clustering, and part-of-speech tagging.

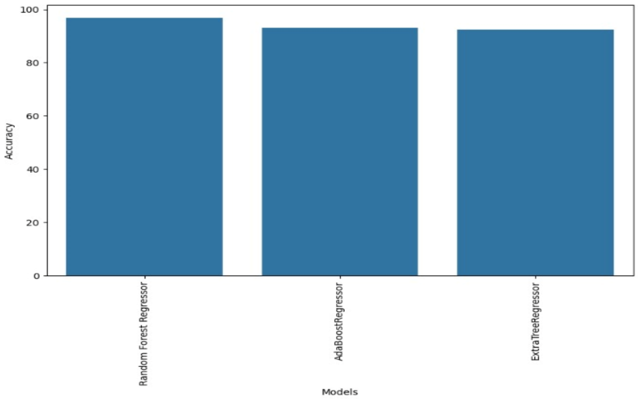
Image Analysis: Image analysisis an application of unsupervised learning. It involves tasks such as image segmentation, object detection, and image pattern recognition.



	Heart_Rate	Skin_Conductivity	Hours_Worked	Stress_Level	Emails_Sent	Meetings_Attended
2	87	5.56	5	28	31	6
3	74	5.09	5	25	42	3
4	70	4.58	9	26	28	4
5	97	5.1	7	30	37	5
6	88	5.23	8	29	35	6
7	60	5.2	7	21	31	6
8	79	5.54	7	26	25	6
9	68	3.18	8	22	30	1
10	64	4.95	10	21	30	7
11	74	5.24	10	25	29	1
12	71	4	9	24	30	2
13	84	6.67	7	27	26	6
14	77	5.16	10	26	27	7
15	71	6.56	10	25	34	2
16	74	4.21	9	25	34	2
17	73	4.03	7	24	31	3
18	84	5.27	10	27	25	3
19	67	3.32	9	23	34	3
20	71	5.71	7	21	26	5
21	61	5.1	10	22	32	4
22	44	6.02	5	17	32	2
23	70	5.7	10	27	37	4
24	78	4.58	8	26	32	6
25	67	3.9	11	20	19	1

RESULTS AND DISCUSSION

The above figure shows that the dataset having the values where we have considered the six attributes based on this attributes we consider the values. It includes heart rate, skin conductivity, number of emails sent, number of hours worked and number of meeting attended.

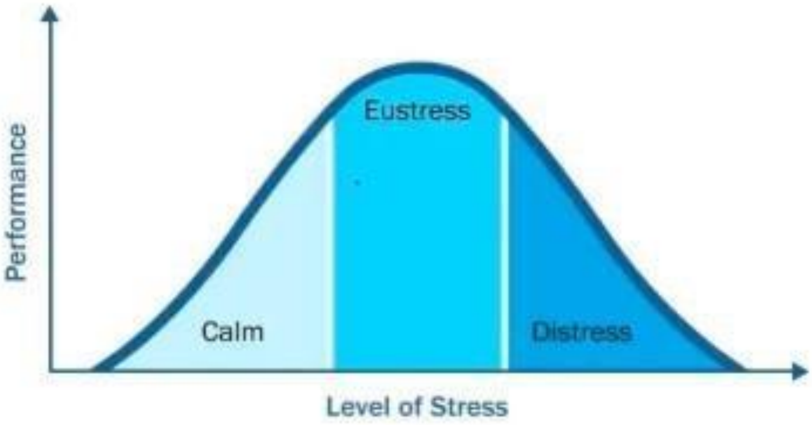


As we considered three different algorithms that are Random forest algorithm, Ada-boost algorithm and Extra trees algorithm, the comparison graph between these algorithms are given above.

S.no.	Algorithm	% of Training data	% of Testing data	Accuracy
1.	Random Forest Algorithm	80	20	99.376%
		70	30	99.376%
		60	40	99.88%
2.	Ada-Boost Algorithm	80	20	92.5%
		70	30	98.5%.
		60	40	92.5%
3.	Extra-Trees Algorithm	80	20	99.5%.
		70	30	97.95%
		60	40	97.95%

The comparison table between three different algorithms based on dataset and accuracy are given above. These comparisons suggest that the Random Forest algorithm generally out performs other machine learning algorithms like SVM and Logistic Regression for stress detection.

The above figure shows the final output of a person stress levels based on given data. The range of stress levels is 0 to 100, with 0 to 25 representing a resting condition, 26 to 50 representing mild stress,51 to75 representing medium stress, and 76 to 100 representing extreme stress.



CONCLUSION

All three algorithms have shown promising results in body stress detection. Random Forest is a robust and efficient algorithm, while Ada-Boost is particularly effective when combined with other algorithms. ExtraTrees is useful when dealing with high-dimensional datasets. The choice of algorithm depends on the specific requirements, such as the type of data, the complexity of the problem, and the desired level of accuracy. In conclusion, Random Forest, Ada-Boost, and Extra Trees are all viable options for body stress detection using different Machine learning algorithms. Each algorithm has its strengths and weaknesses, and the choice of algorithm should be based on the specific requirements of the work.

Future Scope

DNNs have shown remarkable success in stress detection. These networks do not rely on hand-crafted features but instead extract relevant information directly from raw data. Machine learning models can recommend personalized stress-reducing activities based on an individual’s stress patterns. These activities include mindfulness exercises, relaxation techniques, or physical activity, personalized interventions can enhance well-being.

REFERENCES

1. R. Archana , B. M. Devaraju “Stress Detection using Machine Learning Algorithms” in International Journal of Research in Engineering, Science and Management, ISSN : 2581-5792, August 2020.
2. S. A Singh, P. K. Gupta, M. Rajeshwari, and T. Janumala,” Detection of Stress Using Biosensors,” Materials Today: Proceedings 5, 2018, pp. 21003–21010.
3. Fernandes, R. Helawar, R. Lokesh, T. Tari, and A. V. Shahapurkar, “Determination of stress using blood pressure and Galvanic skin response,” in Proc. IEEE Int. Conf. CommunicationandNetworkTechnologies,pp.165–168,201.
4. S. Elzeiny and M. Qaraqe, “Machine Learning Approaches to Automatic Stress Detection: A Review,” in Proc.Int. Conf. Computer Systems and Applications (AICCSA), 2018.
5. Reshma Radheshamjee, and Supriya Kinariwala “Detection and Analysis of Stress using Machine Learning Techniques” International Journal of Engineering and Advanced Technology, pp. 2249 8958, 2019.
6. Pramod Bobade, M.Vani “Stress Detection using Machine Learning and Deep Learning using Multimodal Physiological Data” in Second International Conference on Inventive Research in Computing Applications (ICIRCA), DOI: 10.1109/ICIRCA48905.2020.9183244, July 2020.
7. G. Poornima, B.Ashritha Reddy,M.Kiran Kumar, B.Hrithik Yadav, ”Body Stress Detection Using Machine Learning and IOT Technology” in International Research Journal of Modernization in Engineering Technology and Science, e- ISSN: 2582-5208, 2022.
8. Azhari, Ahmad and Leonel Hernandez," Brain waves feature classification by applying K-Means clustering using single-sensor EEG." International Journal of Advances in Intelligent Informatics 2.3, pp.167-173, Nov 2016.
9. Mehra, Anu, “Study of ECG signals based on gender and heartabnormalities." International Journal of

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

Computational ComplexityandIntelligentAlgorithms,pp.277-291,Feb2020.
10. Barrios-Muriel, Jorge,"A simpleSSA-based de-noising technique to remove ECG interference in EMG signals." BiomedicalSignal Processing and Control, pp. 117-126, Sep 2016.
11. Elgendi M, Mohamed A,Ward R, "Efficient ECG compression and QRS detection for E-health applications," Sci Rep, pp.459-471,Mar. 2017.
12. Kim, Hye-Geum, "Stress and heart rate variability: a meta-analysisandreviewoftheliterature,"Psychiatryinvestigation, pp.235-245,Mar 2018.
13. Behar,JoachimA,"A universal scaling relation for defining power spectral bands in mammalian heart rate variability analysis," Frontiers in physiology, pp. 1001-1005, Aug 2018.
14. Sbrollini, Agnese, et al, "Evaluation of the low-frequency components in surface electromyography," Annual InternationalConferenceoftheIEEEEngineeringinMedicine and Biology Society (EMBC), pp.3622-3626, Aug 2016.
15. Krasteva, Vessela, Irena Jekova, and Ramun Schmid, "Simulating arbitrary electrode reversals in standard 12-lead ECG," Sensors, pp. 20-29, Jul 2019.
16. Balouchestani, Mohammadreza, and Sridhar Krishnan, "Advanced K-means clustering algorithm for large ECG data sets based on a collaboration of compressed sensing theory and K-SVD approach," Signal, Image and Video Processing, pp. 113-120, Oct 2016.
17. Roopa, C.K., B.S. Harish,and SVAruna Kumar, "Classification of ECG Arrhythmia using symbolic dynamics through fuzzy clustering neural network, "Third International Workshop on Pattern Recognition International Society for Optics and Photonics, pp.108-114, Jul 2018.