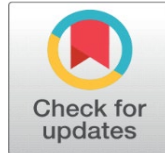# STRESS DETECTION USING MACHINE LEARNING

Lokesh Kr. Sengar [1] ✉ , Sameer Bhati [1] ✉ , Vipul Narayan [1] ✉

[1] Galgotias University, Greater Noida, UP, India

ABSTRACT

We all know that stress today is one of the biggest problems in society, and affects our health indirectly, both physically and mentally. Stress is harmful but if recognized timely can also be prevented and properly handled. This paper provides an overview of the new emerging field of Stress Detection using Machine Learning techniques. A new and intriguing recent stream of research with machine learning — which enables the analysis of vast datasets and recognition of non-linear trends — has been the detection of stress. Our method makes use of a individual's physiological, behavioral and environmental signals and infer their stress levels. For example, Using machine learning algorithms, we can train the models on various features related to stress like age, blood pressure, heart rate of the person to predict whether the person is under stress or not. It may have qualitative characteristics like gender, categories of occupation, or amount of stress. For the classification of human stress level using labeled data, various models can be implemented such as decision tree, random forest, KNN, logistic regression. This abstracts also pointed out the challenges and opportunities in applying ma- chine learning techniques for stress detection. Issues like the need of Huge and Diverse data-sets, moral problems or the chance of model's bias. Since stress is a complex issue, it must be understood in order to help tackle this problem and with the use of technology, software in the case of its use by individuals and communities to manage the stress and with a combination of these two this is a new and emerging area of research and application to machine learning in the world and using technology by persons and individuals makes the stress detection get a good result for a good life until the main task is done.

1. INTRODUCTION

Stress is one of the most common problems across the globe today. Acute and chronic stress has profound consequences in mental health, physical health, and life quality. Stress is leading to a psychological behavior change among individuals it may be a psychosis Jung and Yoon (2017). This study focuses on stress detection with advanced machine learning techniques and how this imperative problem can be handled with advanced methods.

1.1. THE GROWING RELEVANCE OF MENTAL HEALTH ISSUE

Mental health issues have gained prominence in the 21st century. For example, stress has been already identified as an important factor in many psychological and physiological disorders. The implications of stress are

sweeping, affecting not only the individual but also the workplace, families and whole communities. A growing body of research has established the links between stress and a wide range of diseases, including anxiety, depression, heart disease and even weakened immunity, making stress management a high-profile priority in modern health and public health. Moving forward, prioritizing stress as a public health crisis is increasingly crucial. Stress permeates age, gender and socioeconomic lines, making it a universal problem. Even the new global changes like the COVID 19 crises have raised the stress level of the world. This is why the importance of effective, scalable, and skilled stress detection and mitigation techniques could not be clearer Adnan et al. (2012), Nziu and Masu (2019), Rathod and Reddy (2016).

1.2. THE PURPOSE OF PAPER AND ITS CONTRIBUTIONS

This paper embarks on a journey to provide novel insights and methodologies for stress detection. Its central aim is to leverage the capabilities of machine learning, coupled with multimodal data analysis, to craft a comprehensive framework for the early identification of stress. This research is anchored in the belief that by enhancing our ability to detect stress, we can create a more resil- ient and healthier society, where individuals are empowered to confront and manage their stress- ors effectively.

1.3. COMMONLY SOURCES OF STRESS
1.3.1. ENVIRONMENTAL STRESSORS

External stressors arise when individuals are unable to cope with stimuli or situations, leading to stress Adnan et al. (2012). These stressors may include disturbances in the environment, overcrowding, extreme weather conditions, traffic, high crime rates, pollution, and pandemics Adnan et al. (2012) , Nziu and Masu (2019).

1.3.2. SOCIAL STRESSORS

Every person is a part of society and interacts with others in their daily lives. Stress can stem from external factors beyond an individual's control, such as extreme weather, natural disasters, criminal activities, contamination, and death.

1.3.3. PHYSIOLOGICAL STRESSORS

Individuals encounter various stressful situations in their lives as they fulfil different social roles with various people, such as family, friends, colleagues, and partners. Stress can lead to various health issues like weight problems, respiratory complications, diabetes, and asthma Rathod and Reddy (2016).

1.3.4. IMPACT ON STUDENTS

Stress can be especially severe for students, leading to tragic outcomes such as suicides. For instance, a Lancet report in 2012 highlighted a significant number of suicides among individuals aged fifteen to twenty-eight. In 2015, there were 8,934 reported cases of student suicides, and from 2010 to 2015, a total of 39,775 students took their own lives due to stress-related factors Rathod and Reddy (2016) , Bisai and Chaudhary (2017).

1.3.5. TYPES OF STRESS

Stress can be categorized into three types Gjoreski et al. (2016):

1) Acute Stress

This type of stress is short-term, characterized by rapidly developing symptoms but not of ex- tended duration.

2) Episodic Acute Stress

Intense stress occurring during specific periods, such as when students face parental pressure, college exams, or numerous assignments.

3) Chronic Stress

This is a particularly harmful and long-lasting form of stress, persisting for months or even years. It can arise from ongoing situations like relationship issues, family problems, or chronic illnesses.

2. LITERATURE REVIEW

Stress detection, as a field of study, has garnered substantial attention in recent years due to its pivotal role in addressing mental health concerns. This section surveys previous studies on stress identification, summarizing both the strengths and shortcomings of existing approaches and the requirement for novel techniques.

2.1. OVERVIEW OF PREVIOUS RESEARCH IN STRESS DETECTION

Describing methods of stress detection ranges a heterogeneous panorama of data sources and computational techniques. Different types of modalities have been used, such as physiological signals, text data, audio data, etc., and researchers have made notable advancements in this domain. A number of important studies laid the groundwork for more sophisticated stress detection models. We evaluated the model's performance using accuracy, precision, and recall metrics. Our results demonstrate that our machine-learning model can accurately predict with high precision and recall Narayan et al. (2023).

In Khosrowabadi et al. (2011) investigated how more general mechanisms can be used to detect stress and concluded that sensor fatigue makes it unsuitable to assess the overall subjective level of stress over time unless it is carefully monitored and validated over the long term. Each model also takes into account a person's lifestyle habits and medical checkup information for accurate disease prediction Mall et al. (2024). In Narayan et al. (2023) were some of the first researchers to use physiological signals -- heart rate and skin conductance -- to detect the presence of a level of stress in their experimental subjects. Their work was one of the first explorations into using physiology to detect stress, paving the way for later similar research. Moreover, HRV has emerged as a key variable in stress detection research, given that it is a strong measure of the activity of the autonomic nervous system In Mall et al. (2023) Natural language processing (NLP) has been applied in the do- main of text analysis to identify stress in written data (such as social media posts and chat transcripts). In Mall et al. (2024) authors have shown an important links between language and stress, with evidence Narayan et al. (2023) that linguistic features are associated with key elements of emotional states, including stress. These methods showed that stress can be assessed

unobtrusively using text data. Exploration of audio features for stress detection has also attracted attention. In Mall et al. (2023) authors used speech prosody and vocal features to classify stress. And thus audio-based modalities have emerged as a promising avenue, especially in applications that cannot rely on physiological data being available.

2.2. GAPS AND LIMITATIONS IN EXISTING METHODS.

Although great improvements have been achieved, current stress detection solutions also have their drawbacks. Some of the common challenges and shortcomings include: Modality-Dependent Models: Many existing models focus on a single modality (e.g., physiological, text, audio), limiting their ability to capture the full spectrum of stress cues. Multimodal approaches are required to provide a comprehensive assessment.

2.3. GENERALIZABILITY

Some models lack the capacity to adapt to individual differences in stress response and may not perform well in real-world, dynamic settings.

2.4. PRIVACY CONCERNS

Using personal data, such as text messages or physiological signals, raises privacy and ethical issues. Finding a balance between the need for data versus user privacy is a constant struggle.

2.5. TEMPORAL DYNAMICS

Stress is a dynamic process with changing intensities over time. Many models do not adequately address temporal dynamics and rely on static representations of stress Chaturvedi et al. (2024).

2.6. REAL-WORLD APPLICATIONS

While research models may exhibit high accuracy in controlled settings, their practical application in real-world scenarios, such as workplaces, is often underexplored Chaturvedi et al. (2023)

2.7. EMPHASIZING THE NEED FOR NOVEL APPROACHES

The limitations and challenges within the existing literature underline the pressing need for novel stress detection approaches. We address these gaps in this paper in which we propose a multimodal ma- chine learning framework based on novel feature engineering approaches, novel machine learning algorithms, and real-world applicability. Our approach aims to transcend the boundaries of current methodologies by fusing information from diverse sources, adapting to individual variations, and providing practical solutions for stress management. In this current journey, the technical pathway's golden roads show the ultimate metaphor for artificial intelligence (AI) society Mall et al. (2023). Through this literature survey, we have established the foundation for our research and the unique contributions it brings to the field of stress detection. In the subsequent sections, we delve into the specifics of our approach, emphasizing its potential to revolutionize how we detect and address stress in the modern world.

3. DATASET AND DESCRIPTION

This dataset was taken from Kaggle, named as "stress detection". In this dataset we have different rows and columns that represent quantitative measurement such as blood pressure, age, heart rate, stress levels measured on scale. Exploratory data analysis (EDA) is an essential step in under- standing dataset analyse. It involves examining and visualizing data to gain insights and under- standing the dataset Khosrowabadi et al. (2011).
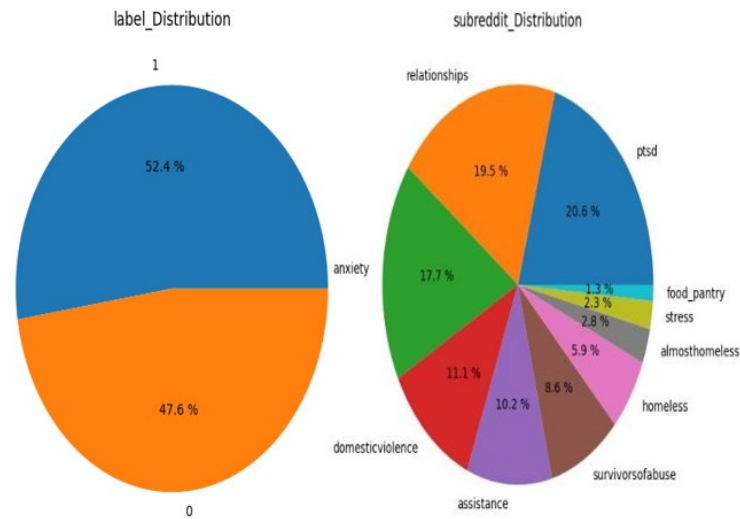
Figure 1

Figure 1 Label and Sub-reddit Distribution

Text processing involves various techniques for handling and analyzing textual data. Some key aspects of text processing include:

3.1. TOKENIZATION

Dividing the text into words or tokens This is a basic step that needs to be performed for many text analysis tasks Narayan et al. (2024).

3.2. TEXT CLEANING

Removing or correcting special characters, punctuation, and unwanted formatting.

3.3. TEXT NORMALIZATION

Standardize the text representation — such as converting all text to lower case.

Figure 2

Figure 2 Dataset

4. EXPERIMENTAL SETUP

We used the python programming language performance parameters, for its implementation:

4.1. RECALL

In machine learning and classification tasks, it is a metric that measures the performance of a model and is particularly useful in binary (two-class) classification problems, for example, when the two classes are referred to as "Positive" and "Negative." Recall (or Sensitivity or True Positive Rate) indicates how effectively a model identifies all relevant instances of the positive class.

The formula for recall is:

Recall = True Positives / (True Positives + False Negatives).

4.2. PRECISION

It is important metric in machine learning and classification task that evaluates the accuracy of positive predictions of a model. It is especially useful in binary classification problems, such as there are 2 classes, generally referred as "positive" and "negative". No of Times True Positive Does up to the No of Times Positive Predictions Done.

The formula for precision is: Precision = TP / (TP + FP).

4.3. ACCURACY

The accuracy is as it represents a basic metric on the quality/degree of predictions done with regard the model. Binary classification is one of the many applications of Logistic Regression. Accuracy is calculated as the ratio of correct predictions growth over the number of total predictions.

The formula for accuracy is: Accuracy = (TP + TN) / (TP + TN + FP + FN)

4.4. F1 SCORE

F1 score is a popular evaluation metric used in machine learning, especially used for binary classification problems. It is an aggregate of precision and recall into a single number and is useful for betting the trade-offs of two given metrics.

The formula to calculate the F1 score is given below:

F1 Score = 2 (Precision * Recall) / (Precision + Recall) There are various strategies that organizations can employ to tackle this issue Mall et al. (2023) Narayan et al. (2023) Mall et al. (2023).

5. CLASSIFICATION ALGORITHM

Machine learning classification approach refers to labelling the data into defined categories or labels via its features or attributes. There are numerous classification algorithms, each with its strengths and weaknesses. Here are some common classification algorithms Chaturvedi et al. (2024) Narayan et al. (2024).

5.1. LOGISTIC REGRESSION

A very interpretable algorithm which models the probability that our instance holds to a specific class Narayan et al. (2024), Mall et al. (2024), Mall et al. (2024).

5.2. DECISION TREES

Build a treelike representation of choices and their potential outcomes. They are easy to under- stand but can be prone to overfitting.

5.3. RANDOM FOREST

This algorithm involves the creation of multiple decision trees, effectively constituting a "forest." This ensemble of decision tree algorithms, often referred to as a "random forest," serves both classification and regression purposes. One notable characteristic of this algorithm is its random selection of the best feature from the available set of features Narayan et al. (2023), Babu et al. (2020).

5.4. K-NEAREST NEIGHBOURS (K-NN)

Assign class instance based on majority population of k nearest neighbours in feature space.

Table 1

Table 1 Comparison of Different Algorithm					
S.no.	**Algorithm**	**Precision**	**Recall**	**Accuracy**	**F1 Score**
1	Random forest	0.72	0.71	0.714789	0.712023
2	Logistic Regression	0.73	0.73	0.733568	0.733494
3	K- Nearest Neighbours	0.68	0.68	0.683099	0.680416
4	Decision Tree	0.61	0.61	0.607981	0.608361

6. CONCLUSION AND FUTURE SCOPE

From the performance Metrics discussed, both Random Forest and Logistic Regression could be good candidates for detecting given dataset. So they give higher precision, recall, accuracy and F1 Score than K-Nearest Neighbour and Decision Tree. But the most appropriate algorithm also depends on use-case need and trade-off between precision and recall. Logistic Regression performs the best

in terms of precision and accuracy, while Random Forest has the most balanced precision and recall and is therefore a good candidate to identify stress levels. It's worth noting that the choice of the algorithm is just one aspect of building an effective stress detection system. Other factors, such as the quality and diversity of the dataset, feature engineering, and model hyperparameter tuning, also play a crucial role in achieving accurate and reliable stress detection results.

In summary, the application of machine learning in stress detection is a promising field with sub- stantial room for growth and improvement. As technology continues to evolve, so does our un- derstanding of stress and its detection. By addressing the challenges and exploring these future avenues, we can work towards more effective stress management and ultimately improve the quality of life for individuals and communities.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Adnan, N., et al. (2012). University Students' Stress Level and Brainwave Balancing index: Comparison between Early and end of study Semester. Research and Development (SCORED), 2012 IEEE Student Conference on. IEEE. https://doi.org/10.1109/SCOReD.2012.6518608

Babu, S. Z., et al. (2020). Abridgement of Business Data Drilling with the Natural Selection and Recasting Breakthrough: Drill data with GA.

Bisai, S., & Chaudhary, R. (2017). Stress Among the Students of an Engineering Institution in India: An empirical analysis. Jindal Journal of Business Research, 6(2), 186–198. https://doi.org/10.1177/2278682117727224

Chaturvedi, P., Daniel, A. K., & Narayan, V. (2023). A Novel Heuristic for Maximizing Lifetime of Target Coverage in Wireless Sensor Networks. Advanced Wireless Communication and Sensor Networks, 227–242. https://doi.org/10.1201/9781003326205-20

Chaturvedi, P., Daniel, A. K., & Narayan, V. (2024). Coverage Prediction for Target Coverage in WSN Using Machine Learning Approaches. Wireless Personal Communications, 137(2), 931–950. https://doi.org/10.1007/s11277-024-11410-x

Gjoreski, M., Gjoreski, H., Lustrek, M., & Gams, M. (2016). Continuous Stress Detection Using a Wrist Device: In the Laboratory and Real Life. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct,1185–1193. https://doi.org/10.1145/2968219.2968306

Jung, Y., & Yoon, Y. I. (2017). Multi-level Assessment Model for Wellness Service Based on Human Mental Stress level. Multimedia Tools and Applications, 76(9), 11305–11317. https://doi.org/10.1007/s11042-016-3444-9

Khosrowabadi, R., Quek, C., Ang, K. K., Tung, S. W., & Heijnen, M. (2011). A Brain-Computer Interface for Classifying EEG Correlates of Chronic Mental Stress. International Joint Conference on Neural Networks (IJCNN),757–762. https://doi.org/10.1109/IJCNN.2011.6033297

Mall, P. K., et al. (2023). A Comprehensive Review of Deep Neural Networks for Medical Image Processing: Recent developments and Future Opportunities. Healthcare Analytics, 4, 100216. https://doi.org/10.1016/j.health.2023.100216

Mall, P. K., et al. (2023). RAnk-Based Two-Stage Semi-Supervised Deep Learning Model for X-ray Images Classification.

Mall, P. K., et al. (2023). Rank-based two-stage semi-supervised deep Learning Model for X-ray Images classification: An Approach Toward Tagging Unlabeled Medical Dataset. Journal of Scientific & Industrial Research (JSIR), 82(08), 818–830. https://doi.org/10.56042/jsir.v82i08.3396

Mall, P. K., et al. (2024). Optimizing Heart Attack Prediction Through OHE2LM: A Hybrid Modelling Strategy. Journal of Electrical Systems, 20(1). https://doi.org/10.52783/jes.665

Mall, P. K., et al. (2024). Self-Attentive CNN+ BERT: An Approach for Analysis of Sentiment on Movie Reviews Using Word Embedding. International Journal of Intelligent Systems and Applications in Engineering, 12(12s), 612–623.

Narayan, V., Daniel, A. K., & Chaturvedi, P. (2023). E-FEERP: Enhanced Fuzzy-Based Energy-Efficient Routing Protocol for Wireless Sensor Network. Wireless Personal Communications, 131(1), 371–398. https://doi.org/10.1007/s11277-023-10434-z

Narayan, V., et al. (2023). A Comprehensive Review of Various Approaches for Medical Image Segmentation and Disease Prediction. Wireless Personal Communications, 132 (3), 1819–1848. https://doi.org/10.1007/s11277-023-10682-z

Narayan, V., et al. (2023). Extracting Business Methodology: Using Artificial Intelligence-Based Method. Semantic Intelligent Computing and Applications, 16, 123. https://doi.org/10.1515/9783110781663-007

Narayan, V., et al. (2024). A Comparison Between Nonlinear Mapping And High-Resolution Image. Computational Intelligence in the Industry 4.0, 153–160. https://doi.org/10.1201/9781003479031-9

Narayan, V., et al. (2024). A Theoretical Analysis of Simple Retrieval Engine. Computational Intelligence in the Industry 4.0, 240–248. https://doi.org/10.1201/9781003479031-13

Nziu, P. K., & Masu, L. M. (2019). Formulae for Predicting Stress Concentration Factors in Flat Plates and Cylindrical Pressure Vessels with Holes: A Review. International Journal of Mechanical and Production Engineering Research and Development.

Rathod, C. H. A. N. D. A. R., & Reddy, G. K. (2016). Experimental Investigation of Angular Distortion and Transverse Shrinkage in CO2 arc Welding Process. International Journal of Mechanical Engineering, 5 (4), 21–28.