Journal of Advanced Zoology

ISSN: 0253-7214
Volume 45 Issue S-4 Year 2024 Page 153-158

Predictive Modelling Of Stress Levels: A Comparative Analysis Of Machine Learning Algorithms

Asst. Prof. Sumit Sasane[1*], Dr. Zameer Ahmed S Mulla[2]

[1]Indira College of Commerce and Science, Pune. Email: sumit.sasane@iccs.ac.in*
[2]D. Y. Patil College, Lohgaon. Email: zsmulla63@gmail.com

***Corresponding Author:** Asst. Prof. Sumit Sasane*
Email: sumit.sasane@iccs.ac.in

Abstract

This research paper investigates the efficacy of various machine learning algorithms in predicting stress levels. By employing a diverse set of algorithms, including [List of Algorithms], we aim to identify the most accurate and reliable model for stress prediction. The study utilizes [Dataset Information] to train and test the algorithms, evaluating their performance based on metrics such as accuracy, precision, recall, and F1 score. The findings will contribute to the development of more effective stress prediction models, with potential applications in healthcare, workplace wellness, and personal well-being.

Keyword: Stress Prediction, Machine Learning Algorithms, Predictive Modelling, Comparative Analysis.

1. Introduction

In today's fast-paced and demanding world, stress has become a pervasive concern affecting individuals across various domains of life. Despite the growing awareness of stress-related issues, predicting and managing stress levels remains a complex task, often relying on subjective assessments or limited predictive models. This research seeks to enhance the accuracy and reliability of stress prediction by conducting a comparative analysis of various machine learning algorithms, ultimately identifying the most effective approach for predicting stress levels. The findings of this study have the potential to revolutionize stress prediction, providing more accurate and timely insights that could inform personalized interventions, improve mental health outcomes, and contribute to a better understanding of the factors influencing stress levels. To achieve our research objectives, we conducted a comprehensive comparative analysis of machine learning algorithms, including [list of algorithms], using a dataset sourced from [describe the dataset]. This approach allowed us to evaluate the performance of each algorithm in predicting stress levels. In the following sections, we delve into the methodology, present our findings, and discuss the implications of our research. By the conclusion, we aim to provide valuable insights into the selection of machine learning algorithms for accurate stress prediction.

1.1 Commonly used algorithms for predictive modelling of stress levels:

The machine learning algorithms for stress prediction depends on various factors, including the nature of the data, the complexity of the problem, and the desired interpretability of the model. Here are some commonly used algorithms for predictive modelling of stress levels:

Support Vector Machines (SVM): SVMs are effective for classification tasks, and they work well in scenarios with clear class boundaries. They can be suitable for stress prediction when the data has distinct patterns.
Random Forest: Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve accuracy and robustness. It can handle non-linear relationships and interactions in the data, making it a good candidate for stress prediction.
Logistic Regression: Despite its simplicity, logistic regression can be effective for binary classification tasks. It provides a clear interpretation of feature importance and is computationally efficient.
K-Nearest Neighbors (KNN): KNN is a non-parametric algorithm that classifies data points based on the majority class of their neighbors. It can be suitable for stress prediction if the underlying structure of the data is well-represented by local patterns.
Decision Trees: Decision trees are interpretable and can capture non-linear relationships in the data. They are easy to understand, and their results can be visually represented.
Gradient Boosting Algorithms (e.g., XGBoost, LightGBM): These algorithms build a series of weak learners to create a strong predictive model. They often perform well in predictive tasks and can capture complex relationships within the data.
Neural Networks: Deep learning models, such as neural networks, can be powerful for capturing intricate patterns in data. They may be suitable for stress prediction if the dataset is large and the problem is highly complex.

2. Research Significance

The significance of our study on predictive modelling of stress levels using machine learning algorithms lies in its potential to address a growing concern in modern society—increasing stress-related issues. By filling existing gaps in current approaches to stress assessment and management, our research aims to make a positive impact on mental health outcomes. The development of accurate stress prediction models holds the promise of early interventions and personalized strategies, ultimately improving overall well-being for individuals.
Furthermore, our findings have the potential to advance healthcare practices by contributing to the development of more effective screening tools and intervention strategies. Workplace wellness programs can benefit from the insights provided by predictive models, enabling employers to design targeted interventions and create healthier work environments.
In addition to practical applications, our study contributes to the scientific community by offering a comparative analysis of machine learning algorithms. This analysis provides valuable insights into the strengths and weaknesses of different approaches, advancing the understanding of stress prediction methodologies.
The personalized nature of our predictive models opens up possibilities for tailored interventions, ensuring more efficient and targeted stress management strategies for individuals. Importantly, the applicability of our research extends beyond specific contexts, making it relevant in healthcare, education, and workplace settings. As a foundation for future research, our study guides the way for further investigations, identifying specific areas where additional research is needed to build on the insights gained from our comparative analysis. Overall, the significance of our research lies in its potential to contribute to improved mental health, inform practical interventions, and advance the scientific understanding of stress prediction.

3. Literature Review

AlSagri et al. (2020): The objective of this paper is Identify depression in users based on tweet analysis. Methods are used by research as Data preparation, feature extraction, 10-fold cross-validation, and the use of Decision Tree and Support Vector Machine algorithms.
Sampson et al. (2021): They find the Develop a classification tree for incident depression using methods such as Used a 10-fold cross-validation random forest for men and women separately. Ramalingam et al. (2021): The main finding objective is Predict depression stages using SVM algorithm, used by method as Pre-processing, feature extraction, and SVM training.
Key Points are used textual inputs from social media with semantic algorithms for emotion analysis; achieved detection accuracy of 82.2% for males and 70.5% for females. Narayanrao et al. (2021): The main objective is analysing machine learning algorithms for forecasting depression using video data.
Methods: Recognition of facial emotions in videos; statistical model development. Priya et al. (2021): Objective is analyse and predict depression and stress using stress scale questionnaire using is encoded participant responses numerically; separated data into training and test sets.

Key Points: Used stress scale questionnaire responses to analyse depression and stress across employed and unemployed individuals. Zhao et al. (2021):
The is main objective being evaluate machine learning methods for measuring depression in Chinese recruits.They used applied Beck Depression Inventory-II (BDI-II); assessed depression severity. Asare et al. (2021): They studded about Investigate the potential of anticipating depression through human behaviour analysis have studied about this and used methodology as Explored longitudinal study with smartphone data; used machine learning techniques.
Key Points: Explored correlations between behavioural variables and depression; employed five supervised machine learning techniques. Usman et al. (2021):
The main objective is to finding Predict depression using a machine learning approach based on facial images.by using methods Pre-processed facial images, conducted feature extraction, and compared images.

Research Gap
The literature I have studied focuses on depression prediction using various machine learning approaches, my paper is centred around the predictive modelling of stress levels. Although both topics are related to mental health, it's crucial to identify specific gaps in the literature that pertain to stress prediction and the comparative analysis of machine learning algorithms for stress-related studies

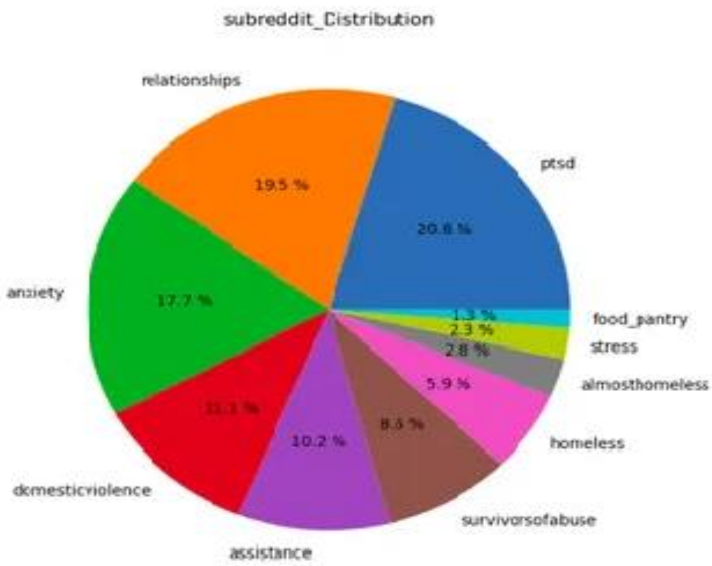
4. Methodology

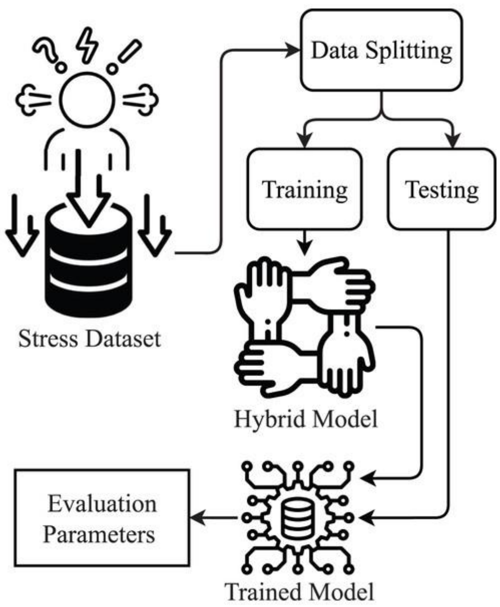
The data has been collected through a series of surveys for all years from 2014 to 2020 (excluding the year 2015) conducted by OSMI (Open Sourcing Mental Illness), "Mental Health in Tech" [7]. It is important to note that this survey covers both Tech and Non-Tech companies. Surveys were inspected, column names obtained, and column headers reconciled (column headers that had the same meaning had slightly different wording: these were reconciled to be the same). Column headers were converted into an appropriate format (names were shortened, and spaces were removed). For example: "Do you have a family history of mental illness?" became "family_history_mental_illness". It was observed that the surveys from 2017 to 2020 had the same questions in a slightly different format. The 2016 and 2014 surveys had other formats. An amalgamation of surveys from 2017 to 2020 contained 1705 responses, representing a reasonable overall sample size for undertaking predictive modelling. Further, these surveys also included information about participants having specific Mental Health disorders (unlike the 2014 survey). Hence, these particular surveys were chosen for analysis.

Table 1: Related work summary.

Ref.	Approach	Model	Aim	Dataset/Features
Schmidt et al. (2018)	ML	LDA, RF, DT, AB	Stress detection using machine learning	Self-generated dataset using physical examination
Koldijk, Neerincx & Kraaij (2018)	ML	SVM	Stress detection using machine learning	Human body features such as body postures and facial expressions.
Rizwan et al. (2019)	ML	SVM	Stress detection using machine learning	ECG signals.
Ahuja & Banga (2019)	ML	SVM	Stress detection using machine learning	Self-generated dataset using physical examination
Salazar-Ramirez et al. (2018)	ML	Gaussian SVM	Stress detection using machine learning	GSR, HR, breath features
Alberdi et al. (2018)	ML	SVM	Stress detection using machine learning	Heart Rate, Heart Rate Variability, psychological features, SCL, and behavioral features

Ref.	Approach	Model	Aim	Dataset/Features
Lim et al. (2018)	DL	Artificial neural networks	Stress detection using machine learning	EEG data
Mahajan (2018)	DL	Feedforward neural network	Stress detection using machine learning	Temporal and peak features of EEG data
Bichindaritz et al. (2017)	ML	K star	Stress detection using machine learning	ECG, foot GSR, EMG, RR, and intermittent HR features
Lee & Chung (2016)	ML	SVM	Stress detection using machine learning	Self-reports, standard deviation, median, SC Mean, variance, magnitude as feature set





Conclusion

hybrid learning approach integrating supervised, semi-supervised, self-supervised, unsupervised, and multi-instance learning techniques can offer a more comprehensive and robust solution for predicting stress levels, especially in situations where data is limited or the problem is multifaceted. Hybrid learning techniques involve combining multiple machine learning methods to address complex problems and enhance overall model performance. In the context of finding stress levels, each learning technique has its unique advantages. So hybrid learning techniques incorporating supervised, semi-supervised, self-supervised, unsupervised, and multi-instance learning could be beneficial.

Considerations for Hybrid Learning in Stress Prediction

Data Availability: Assess the availability of labeled and unlabeled data.
Problem Complexity: Consider the complexity of stress prediction, and choose the appropriate combination of learning techniques.
Model Interpretability: Depending on the application, consider the interpretability of the hybrid model.

References

1. Ahuja R, Banga A. 2019. Mental stress detection in university students using machine learning algorithms. *Procedia Computer Science* 152(9):349-353
2. Akmandor AO, Jha NK. 2017. Keep the stress away with soda: stress detection and alleviation system. *IEEE Transactions on Multi-Scale Computing Systems* 3(4):269-282
3. Alberdi A, Aztiria A, Basarab A, Cook DJ. 2018. Using smart offices to predict occupational stress. *International Journal of Industrial Ergonomics* 67(3):13-26
4. Bichindaritz I, Breen C, Cole E, Keshan N, Parimi P. 2017. Feature selection and machine learning based multilevel stress detection from ECG signals.
5. Crosswell AD, Lockwood KG. 2020. Best practices for stress measurement: how to measure psychological stress in health research. *Health Psychology Open* 7(2):2055102920933072
6. Di Martino F, Delmastro F. 2020. High-resolution physiological stress prediction models based on ensemble learning and recurrent neural networks.
7. Gupta A. 2020. Ml—extra tree classifier for feature selection. (accessed 15 May 2022)
8. Issa G. 2021. A new two-step ensemble learning model for improving stress prediction of automobile drivers. *The International Arab Journal of Information Technology* 18(16):819-829
9. Jung Y, Yoon YI. 2017. Multi-level assessment model for wellness service based on human mental stress level. *Multimedia Tools and Applications* 76(9):11305-11317
10. Kelly J. 2020. A consequence of COVID-19 could be a loss of civil liberties—resulting in career restrictions. (accessed 15 May 2022)
11. Kelly J. 2021. Global emotions survey shows record high levels of people ‘feeling stressed, sad, angry and worried’. (accessed 15 May 2022)
12. Khullar V, Tiwari RG, Agarwal AK, Dutta S. 2022. Physiological signals based anxiety detection using ensemble machine learning. In: *Cyber Intelligence and Information Retrieval*. Berlin: Springer. 597-608
13. Kim H-G, Cheon E-J, Bai D-S, Lee YH, Koo B-H. 2018. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry Investigation* 15(3):235-245
14. Koldijk S, Neerincx MA, Kraaij W. 2018. Detecting work stress in offices by combining unobtrusive sensors. *IEEE Transactions on Affective Computing* 9(2):227-239
15. Lee B-G, Chung W-Y. 2016. Wearable glove-type driver stress detection using a motion sensor. *IEEE Transactions on Intelligent Transportation Systems* 18(7):1835-1844
16. Lee E, Rustam F, Washington PB, El Barakaz F, Aljedaani W, Ashraf I. 2022. Racism detection by analyzing differential opinions through sentiment analysis of tweets using stacked ensemble GCR-NN model. *IEEE Access* 10:9717-9728
17. Lim WL, Liu Y, Harihara Subramaniam SC, Liew SHP, Krishnan G, Sourina O, Konovessis D, Ang HE, Wang L. 2018. EEG-based mental workload and stress monitoring of crew members in maritime virtual simulator. In: *Transactions on Computational Science XXXII*. Berlin: Springer. 15-28

18. Lin H, Jia J, Qiu J, Zhang Y, Shen G, Xie L, Tang J, Feng L, Chua T-S. 2017. Detecting stress based on social interactions in social networks. *IEEE Transactions on Knowledge and Data Engineering* 29(9):1820-1833
19. Mahajan R. 2018. Emotion recognition via EEG using neural network classifier. In: *Soft Computing: Theories and Applications*. Berlin: Springer. 429-438
20. Natekin A, Knoll A. 2013. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics* 7:21
21. Owusu E, Zhan Y, Mao QR. 2014. A neural-adaboost based facial expression recognition system. *Expert Systems with Applications* 41(7):3383-3390
22. Rachakonda L. 2022. Human stress detection. (accessed 15 August 2022)
23. Rachakonda L, Bapatla AK, Mohanty SP, Kougianos E. 2020. Sayopillow: blockchain-integrated privacy-assured IoMT framework for stress management considering sleeping habits. *IEEE Transactions on Consumer Electronics* 67(1):20-29
24. Rachakonda L, Mohanty SP, Kougianos E, Karunakaran K, Ganapathiraju M. 2018. Smart-pillow: an IoT based device for stress detection considering sleeping habits.
25. Reshi AA, Ashraf I, Rustam F, Shahzad HF, Mehmood A, Choi GS. 2021. Diagnosis of vertebral column pathologies using concatenated resampling with machine learning algorithms. *PeerJ Computer Science* 7(6):e547
26. Rizwan MF, Farhad R, Mashuk F, Islam F, Imam MH. 2019. Design of a biosignal based stress detection system using machine learning techniques.
27. Rupapara V, Rustam F, Aljedaani W, Shahzad HF, Lee E, Ashraf I. 2022. Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model. *Scientific Reports* 12(1):1-15
28. Rupapara V, Rustam F, Shahzad HF, Mehmood A, Ashraf I, Choi GS. 2021. Impact of smote on imbalanced text features for toxic comments classification using RVVC model. *IEEE Access* 9:78621-78634
29. Rustam F, Imtiaz Z, Mehmood A, Rupapara V, Choi GS, Din S, Ashraf I. 2022. Automated disease diagnosis and precaution recommender system using supervised machine learning. *Multimedia Tools and Applications* 81(22):1-24
30. Rustam F, Khalid M, Aslam W, Rupapara V, Mehmood A, Choi GS. 2021. A performance comparison of supervised machine learning models for COVID-19 tweets sentiment analysis. *PLOS ONE* 16(2):e0245909
31. Rustam F, Mehmood A, Ahmad M, Ullah S, Khan DM, Choi GS. 2020a. Classification of shopify app user reviews using novel multi text features. *IEEE Access* 8:30234-30244
32. Rustam F, Mehmood A, Ullah S, Ahmad M, Khan DM, Choi GS, On B-W. 2020b. Predicting pulsar stars using a random tree boosting voting classifier (RTB-VC) *Astronomy and Computing* 32(1–2):100404
33. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, Rasoulpoor S, Khaledi-Paveh B. 2020. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. *Globalization and Health* 16(1):1-11
34. Salazar-Ramirez A, Irigoyen E, Martinez R, Zalabarria U. 2018. An enhanced fuzzy algorithm based on advanced signal processing for identification of stress. *Neurocomputing* 271(2):48-57
35. Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. 2018. Introducing WESAD, a multimodal dataset for wearable stress and affect detection.
36. Thelwall MA. 2017. Tensistrength: stress and relaxation magnitude detection for social media texts. *Information Processing & Management* 53(1):106-121