The International Journal of Interdisciplinary Organizational Studies
ISSN: 2324-7649 (Print), ISSN: 2324-7657 (Online)
Volume 19, Issue 1, January-June, 2024

MENTAL HEALTH PREDICTION FOR EMPLOYEES USING MACHINE LEARNING

N.T.Renukadevi
Department of Computer Technology-UG, Kongu Engineering College Erode,India
renuka.ctug@kongu.edu

R.Sridhar
Department of Computer Technology-UG, Kongu engineering college Erode, India
Sridharr.21bir@kongu.edu

S.R. Akash Kumar
Department of Computer Technology-UG, Kongu Engineering College Erode, India
akashkumarsr.21bir@kongu.edu

S.Vignesh
Department of Computer Technology-UG, Kongu engineering college Erode,India
vigneshs.21bir@kongu.edu

Abstract— In today's workplace, addressing mental health challenges is of paramount importance. This project delves into the realm of mental health prediction by harnessing the power of Machine Learning (ML). Utilizing data from a 2014 survey that gauges attitudes toward mental health in the tech industry, the system's primary focus lies in the accurate prediction of mental health consequences within the workplace. Through the implementation of diverse ML algorithms, such as Support Vector Classifier (SVC), Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbours (k-NN), the project assesses the effectiveness of these models in identifying potential mental health risks among employees. With results indicating accuracy rates of 83.59%for SVC, 82.53% for RFC, 74.60% for DT, and 83.33% for k-NN, it becomes evident that machine learning can play a vital role in shaping a more supportive work environment by predicting mental health outcomes. Furthermore, this work paves the way for future enhancements and underscores the significance of algorithm selection and interpretability within the real-world context of organizational mental health support.

Keywords— Mental health prediction, artificial intelligence, machine learning, workplace, classification accuracy.

I. INTRODUCTION

The importance of mental health is becoming more widely acknowledged in the quickly changing workplace, which is why a system centred on mental health prediction was developed. This initiative attempts to proactively identify and solve mental health issues by leveraging data from a 2014 Tech sector survey and the capabilities of machine learning. The system uses a variety of ML methods, such as RF, SVC, KNN, and DT to properly anticipate the effects of mental health on the workplace. This predictive accuracy is a significant advancement in the continuous quest of mental health awareness and support inside companies, providing a basis for creating supportive workplaces and enhancing employee well-being.

Within Artificial Intelligence (AI), ML is a dynamic discipline that enables systems to learn from data automatically and perform better on tasks without explicit programming. It includes a variety of methods and algorithms that let computers analyze, comprehend, and draw patterns from massive information. For applications like data-driven decision-making, prediction, and classification, it's a priceless tool. Machine learning is a key factor in innovation and automation, with applications ranging from image recognition and natural language processing to healthcare and finance. It has the potential to completely change how we approach complex issues and derive insightful information from the ever-increasing amounts of data available in the digital age.

The actual site where employees perform their job duties is referred to as the workplace. It encompasses the workspace or office where duties are carried out and communications with supervisors and co-workers take place. It can also refer to the larger workplace, which includes elements like corporate culture, rules, and general working conditions.

The ability of a predictive model or system to accurately foresee or anticipate future events or outcomes is measured by its predictive accuracy. By contrasting the forecasts with the actual observed data, it evaluates the model's propensity to make correct predictions. A greater predictive accuracy denotes a tight match between the model's predictions and reality, whereas a lower accuracy denotes less trustworthy forecasts.

In the context of the workplace, a supportive environment is one in which staff members have access to the tools, motivation, and emotional support they need to do their jobs successfully and take care of themselves. It includes things like a culture that puts employee mental and emotional wellness first, open communication, management support, and access to mental health services. The goal of a supportive atmosphere at work is to enable staff members to realize their full potential while preserving their mental and emotional health.

The remaining sections of the article are organized into four parts. The first section, encompasses an in-depth literature review on stress detection, whiles the second section, details the dataset employed in the study. The third section, explains about the methodologies proposed in this research work followed by Results and Discussion section which showcases the outcomes of machine learning and deep learning models applied to stress detection. Lastly, the article concludes by summarizing the key findings and implications.

II. LITERATUREREVIEW

The research work presented by (Piya Majumdar & Sahu, 2020) has mentioned that the globe has entered unknown territory because to COVID-19. The economy has collapsed, several nations are under lockdown, and nearly everyone is terrified of serious repercussions. Numerous difficulties arise from the extraordinary shifts that occurred so fast as a result of the pandemic and confinement to one's house to achieve social distance and reduce infection risk. Among these are weakened healths, happiness, and sleep due to interruptions in everyday life; anxiety, worry, loneliness; increased stress from job and family; and too much screen time.

This research (Ciccarelli et al., 2023) focuses on the prevalent issue of stress-induced depression, which has become a significant concern not only in academic studies but also in people's social lives. The growing awareness of the impact of stress on various aspects such as quality of life, employment, social interactions, and health has led to an increased demand for tools and software capable of assessing and managing stress in real-time. Recent advancements in science and technology have spurred the development of novel techniques and strategies, particularly in the context of stress assessment in the workplace, a major source of stress in contemporary society.

In their research, (Giannakakis et al., 2019) explore the impact of psychological stress on the human body through the analysis of bio-signals. The central objective of this psycho-physiological study is to establish reliable indices derived from bio-signals that shed light on the physiological mechanisms underlying the stress response. The motivation behind this investigation arises from the existing literature's lack of comprehensive guidelines detailing the correlation between the properties of utilized bio-signals and the corresponding behavioral outcomes during stress. This paper extensively examines the influence of stress on diverse physiological responses, placing particular emphasis on assessing the effectiveness, resilience, and consistency of bio-signal data properties within the current state-of-the-art for stress detection. The overarching goal of this article is to provide a thorough overview of the patterns of bio-signals triggered by stressful conditions, offering practical and robust guidelines for more efficient stress detection.

In their research, (Gedam & Paul, 2021) underscore the significance of early stress detection for mitigating its potential impact, emphasizing the feasibility of achieving this through continuous stress monitoring. Wearable devices play a pivotal role in enabling real-time and ongoing data collection, facilitating personalized stress monitoring. This article provides a comprehensive review centered on stress detection using wearable sensors and the application of machine learning techniques. The investigation explores stress detection methods that utilize sensory devices like wearable sensors, Electrocardiogram (ECG), Electroencephalography (EEG), and Photoplethysmography (PPG) across various environments, including driving, studying, and work settings. Each study's stressors, methodologies, outcomes, advantages, limitations, and challenges are systematically examined, offering insights and proposing directions for future research. In conclusion, the paper suggests a multimodal stress detection system that integrates wearable sensor-based deep learning technology.

(Can et al., 2019) draw attention to the limitations in current stress detection research, particularly concerning everyday life scenarios involving smartphones and mobile devices. Despite notable advancements in stress detection within controlled laboratory settings, there is a notable dearth of studies examining stress detection in real-world, daily situations. The survey systematically categorizes and analyses existing works based on the physiological indicators used and the specific environments targeted, such as offices, campuses, cars, and various daily conditions.

(Dai et al., 2021) utilized machine learning techniques by employing physiological signals to construct stress prediction models. Following thorough data processing using signal processing techniques, two types of models were formulated: objective stress models utilizing stressor tasks as identifiers and subjective stress models employing participants' responses to each task as labels for stressful tasks. Support Vector Machines (SVM) demonstrated superior performance for objective stress models, achieving an AUROC value of 0.790 and an F-1 score of 0.623. SVM also exhibited the best performance for subjective stress models, with an AUROC value of 0.726 and an F-1 score of 0.520. The model's performance further improved to an AUROC value of 0.775 and an F-1 score of 0.599 with the customized threshold model. Enhancing the effectiveness of personalized threshold-based models opens new avenues for the advancement of stress prediction models.

Early prediction of stress levels is crucial for mitigating its effects and preventing serious health issues associated with this mental state. Automated systems are necessary for accurate stress level prediction. This study introduces an approach that efficiently detects stress using machine learning techniques. The proposed hybrid model (HB) combines gradient boosting machine (GBM) and random forest (RF) through soft voting criteria, utilizing each model's prediction probability for the final outcome. Notably, the proposed model achieves a significant 100% accuracy compared to state-of-the-art approaches. To further demonstrate the approach's significance, 10-fold cross-validation is conducted, with the proposed HB model outperforming others with a mean accuracy of 1.00 and a standard deviation of +/−0.00. Additionally, a statistical T-test is employed to underscore the significance of the proposed approach when compared to alternative methods (Ding, Cheng et al., n.d.)

A comprehensive examination of prior research on stress detection utilizing machine learning algorithms were presented in (Kene & Thakare, 2021). It introduces a stress level classification framework utilizing the PhysioBank dataset for analysing stress levels. Feature selection and extraction were carried out through statistical analysis, revealing the successful implementation of stress level classification using the proposed gradient boost algorithm. The evaluation results demonstrated the effectiveness of the model with accuracy (83.33%), specificity (75%), sensitivity (75%), positive predictive value (90%), and negative predictive value (90%), and error rate (16.66%), F1_Score (83.33%), and recall (75%). Comparative analysis with other machine learning algorithms, including KNN, Random Forest, and support vector machine, indicated the superior performance of the proposed gradient boost algorithm. The findings highlight the efficacy of the proposed model in accurately classifying stress level predictions.

A research work focuses on predicting anxiety, depression, and stress levels using machine learning algorithms. Data were gathered from employed and unemployed individuals across diverse cultures and communities through the Depression, Anxiety, and Stress Scale questionnaire (DASS 21). The predictions for anxiety, depression, and stress, categorized into five severity levels, were made using five different highly accurate machine learning algorithms tailored for predicting psychological problems. During the application of these methods, it was observed that the confusion matrix indicated imbalances in the classes. To address this, the f1 score measure was introduced, aiding in the identification of the Random Forest classifier as the most accurate model among the five algorithms tested. Additionally, the specificity parameter highlighted the algorithms' heightened sensitivity to negative results. (Priya et al., 2020).

When an individual experiences stress, various bio-signals which exhibit discernible patterns that can be utilized for stress level identification. This paper employs a dataset obtained through an Internet of Things (IoT) sensor, providing real-life insights into an individual's mental health in stressful situations. Data from sensors like the Galvanic Skin Response Sensor (GSR) and Electrocardiogram (ECG) were collected to establish patterns for stress detection. The dataset is then categorized using Multilayer Perceptron (MLP), Decision Tree (DT), K-Nearest Neighbour (KNN), Support Vector Machine (SVM), and Deep Learning algorithms (DL). Performance evaluation metrics such as accuracy, precision, recall, and F1-Score are employed to assess the effectiveness of the data analysis. Ultimately, Decision Tree (DT) emerges as the top-performing classifier, achieving an accuracy of 95%, precision of 96%, recall of 96%, and F1-score of 96% among all machine learning classifiers. (Zainudin et al., 2021)

An overview of Machine Learning (ML) methodologies are applied in the identification and examination of mental stress and disorders. The analysis reveals that Support Vector Machines (SVM), Neural Networks (NN), and Random Forest (RF) models consistently demonstrate superior performance in terms of accuracy and resilience compared to other ML algorithms. Physiological indicators, such as heart rate and skin response, are frequently utilized as stress predictors. The review underscores the significance of employing dimensionality reduction techniques prior to ML algorithm training. The results highlight gaps in research and propose future avenues, including the need for interpretable models, personalized approaches, exploration in naturalistic settings, and the development of real-time processing capabilities for stress detection and prediction. (Razavi et al., 2023)

(Chung & Teo, 2022) presented aresearch work which addresses the growing prevalence of mental health issues and the increasing demand for effective healthcare solutions, prompting an exploration of machine learning applications in predicting mental health problems. The study presents a recent systematic review of various machine learning approaches in this domain, discussing challenges, limitations, and potential future directions.

In a study proposed by (Vaishnavi et al., 2022), five machine learning techniques were identified and evaluated for their accuracy in detecting mental health issues using various criteria. The selected machine learning techniques included Logistic Regression, K-NN Classifier, Decision Tree Classifier, Random Forest, and Stacking. These techniques were compared through implementation, and the Stacking technique emerged as the most accurate, achieving a prediction accuracy of 81.75%.

An innovative prediction model that combines an optimization algorithms and neural network to identify and rank the most influential factors impacting the mental health of medical professionals is proposed by (Wang et al., 2021). Utilizing stepwise logistic regression, binary bat algorithm, hybrid improved dragonfly algorithm, and our proposed prediction model, we conducted predictions for the mental health of medical workers. The outcomes indicate that our proposed model achieves a prediction accuracy of 92.55%, surpassing the performance of existing algorithms. This approach holds promise for the global prediction of mental health in medical workers and can contribute to devising tailored work plans for this demographic.

An approach to predict mental health disorders in social network were analyzed using ML algorithms like K-NN, SVM and CNN which shows better accuracy based on their tweets. (T. E. Ramya, n.d.) Stress, anxiety level and depression were analyzed to predict psychological issues using ML algorithms. (Pugazharasi et al., 2023). Precision medicine is being involved in taking care of patients to diagnose diseases and medicines (Ponselvakumar et al., 2021)

III. PROPOSED METHODOLOGY

The suggested system is a comprehensive machine learning solution made to anticipate, manage, and deal with mental health issues that arise in the workplace. A number of crucial phases are included in this system, such as feature engineering, data collecting, pre-processing, model selection, and prediction. Its main goal is to forecast the effects on mental health using a variety of machine learning methods, including K-Nearest Neighbors, Support Vector Classifier, Random Forest, and Decision Tree. These methods seek to proactively detect possible mental health hazards among workers by efficiently utilizing data, hence promoting a more resilient and supportive work environment. The system's ultimate goal is to enhance worker wellbeing by treating mental health issues using data-driven treatments and forecasts.

A. DATASET DESCRIPTION

The current approach uses machine learning techniques to analyse stress patterns and identify important elements impacting stress levels, so addressing the common problem of stress illness among IT professionals. Because of changing work cultures and lifestyles, stress is still a major worry in the workplace, even with the efforts made by businesses and sectors to enhance mental health assistance and conditions. This research makes use of information from 2017 The Open Source Mental Illness team conducted mental health survey, gathered from kaggle repository, which included responses from working tech employees. This dataset includes 756 records and this study finds the factors that affect stress levels include gender, family history, and availability of health benefits at work.

B. SUPPORT VECTOR MACHINE

SVM is a highly popular supervised learning method employed for classification and regression tasks, though it predominantly targets classification. Its aim is to find the most effective decision boundary, referred to as a hyper plane that can accurately segregate classes within an n-dimensional space. This is accomplished by identifying support vectors, which are critical data points positioned at the extremes and essential for delineating the hyperplane. It derives from these pivotal support vectors, which are instrumental in determining the decision boundary for classifying future data points. The diagrammatic representation of SVM is shown in Figure 1.

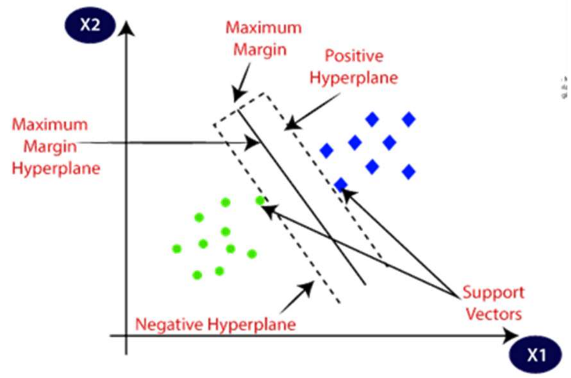

Fig.1. Support Vector Classifier

C. K-NN

KNN is a core supervised learning technique that classifies a new instance by identifying the category most akin to existing categories. It relies on the premise that the new instance shares similarities with the examples already present in the dataset. By storing relevant data and classifying new data points based on similarity, K-NN swiftly assigns them to suitable categories. While predominantly used for classification, K-NN can also be applied to regression tasks. K-NN, being a non-parametric algorithm, operates without presumptions about the underlying data. It's commonly termed a "lazy learner" algorithm due to its characteristic of deferring learning from the training set. The process of K-NN is depicted in Figure 2. Rather than immediately learning from the data, it classifies new instances by comparing them to the existing dataset, rendering it well-suited for real-time classification during training.

Fig.2. K-NN Process

D. DECISION TREE

Decision trees, although primarily applied to classification challenges, can also address regression problems within the realm of supervised learning. This classifier employs a hierarchical tree arrangement, where internal nodes denote dataset attributes, branches indicate decision criteria, and terminal nodes signify outcomes. The tree structure consists of Decision Nodes, which determine choices with various branches, and Leaf Nodes, which depict final results without additional branches. Decisions at the decision nodes are guided by the characteristics of the dataset, serving as a graphical tool that illustrates problem-solving options or decision-making processes based on specific parameters.

E. RFC

Random Forest, a widely recognized supervised learning algorithm, is capable of addressing both classification and regression problems in machine learning. It operates on the principle of ensemble learning, where multiple classifiers are combined to tackle complex issues and enhance overall model performance. Random Forest is a classifier that encompasses numerous decision trees built on different subsets of the provided dataset, utilizing the average to enhance predictive accuracy. Unlike relying on a single decision tree, the random forest predicts outcomes by aggregating majority votes from projections generated by each tree.

IV. EXPERIMENTALRESULTS

The experimental results were evaluated with performance metrics such as Classification Accuracy which is the ratio of number of correct predictions among all the total number of data points. In order to evaluate classification accuracy, confusion matrix is used for every classification algorithms, which is shown in Figure 3 and Figure 4.

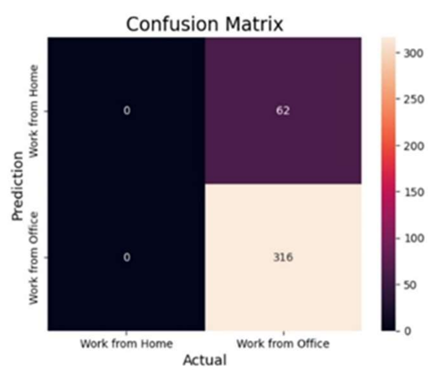

Fig.3. Confusion Matrix of SVC Algorithm

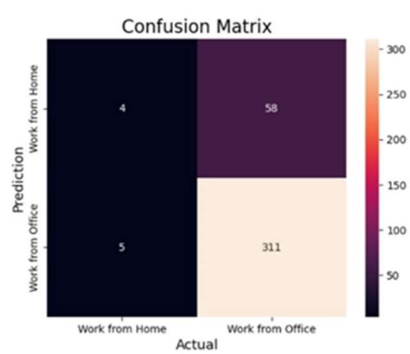

Fig.4. Confusion Matrix of KNN Algorithm

Our machine learning analysis's findings showed that different algorithms used to forecast the effects of mental health issues at work had differing degrees of accuracy. With an accuracy of

83.59%, SVC was the best-performing model; KNN with 83.33%, was next in line. With 82.53% accuracy, RFC outperformed Decision Tree, which came in last with 74.60%. The comparison of classification accuracy of each algorithm is shown in Figure 5. With SVC and KNN showing promising results, these studies highlight the potential of machine learning to help detect and treat mental health issues in the IT field. Nevertheless, when selecting the best method, it is crucial to take into account model interpretability and the particular implementation situation, as these aspects are crucial in practical applications. Additional research, likes as feature engineering and hyper parameter tweaking, might improve the predicted accuracy and provide more in-depth understanding of this important facet of worker wellbeing.

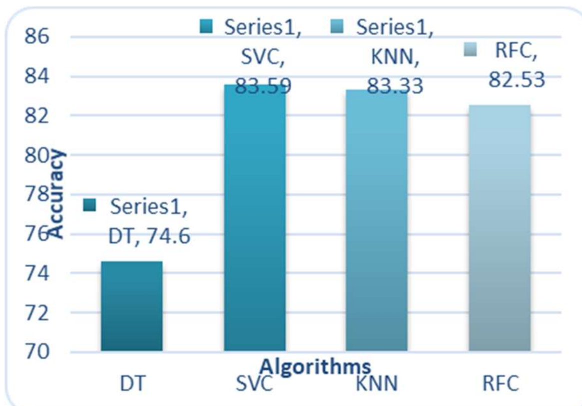

Fig.5. Classification Accuracy of ML Algorithms

V. CONCLUSION

Our machine learning investigation concludes by showing the promise of predictive models in recognizing and resolving workplace mental health issues, especially in the tech sector. The algorithms used—RFC and SVC being the most successful—offer useful resources for companies looking to enhance employee well-being proactively. Future research might concentrate on feature engineering, hyper parameter optimization, and adding more data sources to increase the predictive potential of these models. Furthermore, for real-world applications, a comprehensive strategy that takes into account both interpretability and model correctness is essential. The future use of these cutting-edge technologies may result in happier, more encouraging.

REFERENCES

1. Can, Y. S., Arnrich, B., & Ersoy, C. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. Journal of Biomedical Informatics, 92, 103139. https://doi.org/10.1016/j.jbi.2019.103139

2. Chung, J., & Teo, J. (2022). Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. Applied Computational Intelligence and Soft Computing, 2022, 9970363. https://doi.org/10.1155/2022/9970363

3. Ciccarelli, M., Papetti, A., & Germani, M. (2023). A review of work-related stress detection, assessment, and analysis on-field. Procedia CIRP, 120, 1220–1225. https://doi.org/10.1016/j.procir.2023.09.152

4. Ding, Cheng et al. (n.d.). A systematic hybrid machine learning approach for stress prediction.

5. Gedam, S., & Paul, S. (2021). A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques. IEEE Access, 9, 84045–84066. https://doi.org/10.1109/ACCESS.2021.3085502

6. Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2019). Review on Psychological Stress Detection Using Biosignals. IEEE Transactions on Affective Computing, PP, 1–1. https://doi.org/10.1109/TAFFC.2019.2927337

7. Kene, A., & Thakare, S. (2021). Mental Stress Level Prediction and Classification based on Machine Learning. 2021 Smart Technologies, Communication and Robotics (STCR), 1–7. https://doi.org/10.1109/STCR51658.2021.9588803

8. Piya Majumdar, A. B., & Sahu, S. (2020). COVID-19 pandemic and lockdown: Cause of sleep disruption, depression, somatic pain, and increased screen exposure of office workers and students of India. Chronobiology International, 37(8), 1191–1200. https://doi.org/10.1080/07420528.2020.1786107

9. Ponselvakumar, A. P., Anandamurugan, S., Logeswaran, K., Nivashini, S., Showentharya, S. K., & Jayashree, S. S. (2021). Advancement in Precision Medicine and Recommendation System for Clinical Trials Using Deep Learning Methods. IOP Conference Series: Materials Science and Engineering, 1055(1), 012110. https://doi.org/10.1088/1757-899X/1055/1/012110

10. Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. Procedia Computer Science, 167, 1258–1267. https://doi.org/10.1016/j.procs.2020.03.442

11. Pugazharasi, K., Kalaivani, P., Jayapriya, J., Kadhambari, S., & Mailvizhi, S. (2023). Machine learning algorithms for predicting depression, anxiety and stress in modern life. AIP Conference Proceedings, 2587(1), 090007. https://doi.org/10.1063/5.0150604

12. Razavi, M., Ziyadidegan, S., Jahromi, R., Kazeminasab, S., Janfaza, V., Mahmoudzadeh, A., Baharlouei, E., & Sasangohar, F. (2023). Machine Learning, Deep Learning and Data Preprocessing Techniques for Detection, Prediction, and Monitoring of Stress and Stress-related Mental Disorders: A Scoping Review.

13. T. E. Ramya. (n.d.). An Effective Approach for Mental Health Prediction Using Machine Learning algorithm. 2022, 10(13), 81–84. https://doi.org/10.17577/IJERTCONV10IS13016

14. Vaishnavi, K., Kamath, U. N., Rao, B. A., & Reddy, N. V. S. (2022). Predicting Mental Health Illness using Machine Learning Algorithms. Journal of Physics: Conference Series, 2161(1), 012021. https://doi.org/10.1088/1742-6596/2161/1/012021

15. Wang, X., Li, H., Sun, C., Zhang, X., Wang, T., Dong, C., & Guo, D. (2021). Prediction of Mental Health in Medical Workers During COVID-19 Based on Machine Learning. Frontiers in Public Health, 9. https://doi.org/10.3389/fpubh.2021.697850

Zainudin, Z., Hasan, S., Shamsuddin, S. M., & Argawal, S. (2021). Stress Detection using Machine Learning and Deep Learning. Journal of Physics: Conference Series, 1997(1), 012019. https://doi.org/10.1088/1742-6596/1997/1/012019