Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 2s (2026) | Pages: 166-171

Stress Prediction Using ML Algorithm

Rashmi Adatkar[1], Dharti Jethwa[2], Durvesh Martal[3]

[1,2,3]Electronics and Telecommunication, K.J. Somaiya Institute of Technology, Mumbai,India

[1]rashmi@somaiya.edu, [2]dhartijethwa23@gmail.com, [3]martaldurvesh@gmail.com

Peer Review Information

Type: *Article*
Received: *27 March 2026*
Revised: *12 April 2026*
Accepted: *26 May 2026*
Published: *16 June 2026*

Abstract

In the current scenario, mental stress has become a major issue in society, impacting mental and physical well- being. Early detection of the level of stress is critical in order to intervene and avoid severe mental conditions. In this paper, a novel method of predicting stress levels using the responses obtained in the DASS-21 questionnaire is presented using the machine learning approach. In the proposed system, a Random Forest classifier is used to classify individuals based on different levels of stress severity. Data preprocessing techniques are also used to enhance the accuracy of the model, including label encoding, feature scaling, and handling missing values. In the experiment, it has been shown that the Random Forest model is capable of predicting the levels of stress and the most dominant features in the questionnaire.

Keywords: Stress Prediction; DASS-21 Questionnaire; Random Forest Classifier; Machine Learning; Mental Health Assessment; Feature Importance; Psychological Stress.

How to Cite This Article

Adatkar, R., Jethwa, D., & Martal, D. (2026). Stress prediction using ML algorithm. *Multidisciplinary Journal of Research in Engineering and Technology*, 13(2s), 166–171.

Introduction

Stress is one of the common psychological conditions that people encounter in their life because of various personal, academic, and professional pressures. When people are under stress for a long time, they may develop severe mental health problems like anxiety and depression. Therefore, the detection of stress is critical. Previously, the detection of stress was done manually, which is time-consuming. However, the develop- ment of efficient machine learning techniques has introduced the concept of stress prediction systems.

The Depression Anxiety Stress Scales (DASS-21) is a standard instrument for measuring mental states based on self- reported data. Manual processing of the results of the DASS- 21 may not be possible for large-scale data, whereas machine learning techniques can process the results of the DASS-21 easily and make accurate predictions. Among many other ma- chine learning techniques, Random Forest is a popular method for its reliability, non-linearity, and resistance to overfitting.

In this work, a stress prediction model based on the Ran- dom Forest algorithm is proposed, considering the DASS-21 questionnaire. This system processes the questionnaire, selects the proper preprocessing technique, and predicts the level of stress. Furthermore, the importance of the features is analyzed to understand the impact of the questionnaire in the stress prediction model. The proposed model is expected to offer an efficient solution for the prediction of stress level.

Review Of Literature

Stress Assessment in Academic and Professional Environments

Stress has been recognized as an important factor that impacts the mental health, academic performance, and productivity of students as well as working individuals. Previous research studies that focused on the assessment of stress have used methods such as clinical interview-based evaluation and self- assessment evaluation methods carried out by mental health professionals. Though these methods are effective, they are time-consuming. In addition, it is observed that people often face stress that is not of a clinical nature but is in the initial stages. This is the reason why automated methods of stress assessment are required.

Use of DASS-21 Questionnaire for Stress Measurement

Depression Anxiety Stress Scales, also known as the DASS- 21, is an established measure for determining the stress, anxiety, and depression levels of an individual. Various re- searchers have proven the efficacy of the DASS-21 question- naire for different demographic groups, especially for students. Conventional methods based on the DASS-21 questionnaire rely on the summation of the results obtained. However, the conventional method may fail to consider the intricate relationship between the responses provided for each question in the questionnaire. Various studies have proven that the results obtained from the questionnaire can be enhanced using sophisticated analysis.

Machine Learning Approaches for Stress Prediction

With the advancement of machine learning, researchers have attempted to utilize data-driven techniques to achieve higher accuracy in stress prediction. Techniques such as Naive Bayes, Decision Trees, Support Vector Machines, and Neural Net- works have been used on questionnaire-based data sets. These techniques showed better performance than traditional sta- tistical techniques. Some techniques face problems such as overfitting, noise sensitivity, or lack of interpretability, which makes it difficult to apply these techniques in real-world scenarios.

Random Forest in Psychological Data Analysis

Ensemble learning methods, especially the Random Forest method, have been widely adopted in the field of mental health research due to their robustness and reliability. The Random Forest model is based on the combination of multiple decision trees to reduce variance and achieve better results in terms of generalization. According to various studies, the Random Forest model is highly efficient for high-dimensional ques- tionnaire data and is able to handle non-linear relationships between various psychological measurements. Moreover, the feature importance property of the Random Forest model helps in the identification of significant stress-related features.

Digital and Automated Stress Screening Systems

Modern literature discusses the significance of mental health screening via digital tools, with particular relevance to edu- cational settings. Using automated systems that use machine learning allows one to assess stress immediately, detect at- risk persons quickly, and implement the program en masse without requiring many resources for clinical services. This technology is especially significant in areas where access to such specialists is limited. Scholars state that such tools should be applied only as decision-support systems, not as substitutes for clinicians.

Proposed System

- Data Collection: The first phase is the collection of data, which is done through a standardized stress survey based on the DASS model, focusing on the workplace and academic settings. The data collected includes responses to questions on stress-related issues, which are indicated by emotional tension, tiredness, difficulty in relaxing, and work or study-related pressure. This data is the main input to the stress prediction model.

- Data Pre-processing: In raw data, it is usually observed that there is missing data, inconsistent data, and data related to categories, which cannot be directly handled by machine learning algorithms. To deal with this, pre-processing is done, which includes handling missing data, encoding categorical data, and scaling numerical data.

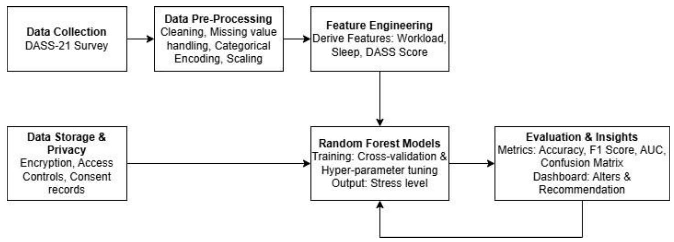

Fig. 1. *Block Diagram of Stress Prediction Model*

- Feature Engineering: Feature engineering is done in order to improve the predic- tive power of the model. Relevant features are extracted, and these include workload, sleep-related features, and aggregated stress levels using the DASS score. This is done in order to capture meaningful features, which play a significant role in predicting the stress levels.

- Data Storage and Privacy: Considering the sensitive nature of mental health-related data, secure data storage is achieved. Encryption, access control, and user consent are also used to ensure the privacy and confidentiality of the data collected from the participants.

- Random Forest Model Training: A Random Forest classifier is employed as the core predic- tion model, which is chosen for its robustness, particularly for high-dimensional data. Cross-validation is employed for the training of the model, along with hyperparameter optimization for the number of trees and depth. The prediction model is then used to predict the levels of stress, including normal, mild, moderate, and severe.

- Evaluation and Insights Generation: The performance of the proposed model is evaluated using various parameters like accuracy, precision, recall, F1-score, and confusion matrix. In addition, feature importance is also used to identify the most dominant features that affect the stress level. It is also possible to add features like dashboards, notifications, and recommendations to the system to help in the identification of stress.

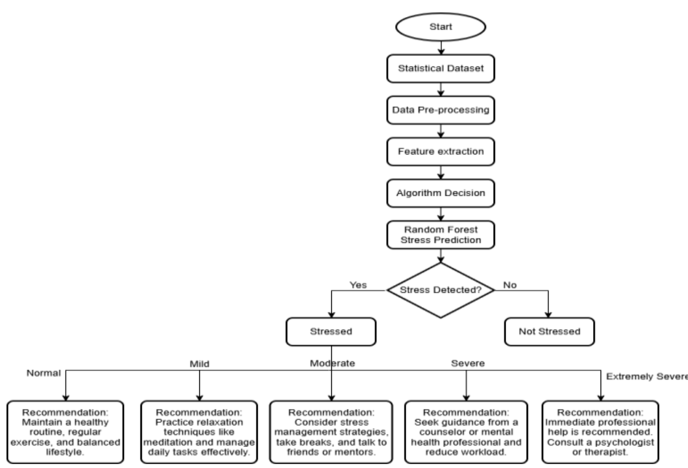

Fig. 2. *Flow Chart of the Model*

The proposed framework adopts a step-by-step workflow for stress detection using the machine learning algorithm. First, the statistical dataset is generated by answering the DASS-21 questionnaires, capturing psychological parameters indicative of stress, anxiety, and depression. Next, the data is preprocessed through missing values imputation, categorical encoding, and feature scaling to guarantee consistent data entry. Features that capture information from the answers are extracted from the dataset and used as inputs in the machine learning model. Random Forest classifier is chosen due to its efficiency in dealing with noise in the dataset. This model predicts if there is stress, followed by a classification into five types: Normal, Mild, Moderate, Severe, and Extremely Severe. Depending on the predicted stress type, specific actions can be recommended, such as engaging in physical exercise to relieve mild stress and receiving psychological treatment for severe and extremely severe stress.

Methods

1. Participants and Setting: The study consisted of participants from the academic pop- ulation and the general population, who volunteered to take the online stress assessment survey. Participation in the study was voluntary, and informed consent was obtained from all the participants before the actual study began. An online survey was used to ensure that participation was easy for the volunteers. Anonymity was ensured for the participants to maintain confidentiality.
2. Instruments: The assessment was made by employing the Depression Anx- iety Stress Scale-21, which is a reliable and well-validated tool for measuring stress levels. The DASS-21 is composed of 21 questions that assess psychological distress through three subscales: Depression, Anxiety, and Stress. In this study, stress dimension results were used to assess stress levels. Each question is rated on a Likert scale, and results are used to categorize respondents into five stress levels: Normal, Mild, Moderate, Severe, and Extremely Severe.
3. Procedure: In the process of collecting the data, a structured online form was created with the 21 questions in the DASS-21 tool. The participants then submitted their responses electronically. After that, the data went through the preprocessing phase, which involved the handling of missing values, scaling, and encoding. Finally, the data was used to predict the level of stress severity using a machine learning model that had been trained previously.
4. Data Analysis: The collected data is then analyzed using various machine learning algorithms through the Python programming lan- guage. The data is then normalized using feature scaling, and the Random Forest Classifier is used to classify the stress severity classification model. The model is then trained and validated using cross-validation and tuning of the model to ensure robustness. The model is then evaluated based on accuracy, confusion matrices, and feature importance metrics.

Results And Discussion

Results

The suggested random forest-based stress prediction model was tested based on the Kaggle data set as well as the real time data set that acted as the main testing dataset. The performance of the model was evaluated using a confusion matrix as well as some other classification measures, including accuracy, precision, recall, and the F1 measure.

Classification Performance of Kaggle Dataset

```
Classification Report:
                  precision    recall  f1-score   support

Extremely Severe       0.00      0.00      0.00         1
            Mild       0.00      0.00      0.00         4
        Moderate       0.17      0.20      0.18         5
          Normal       1.00      0.60      0.75         5
          Severe       0.38      0.60      0.46         5

        accuracy                           0.35        20
       macro avg       0.31      0.28      0.28        20
    weighted avg       0.39      0.35      0.35        20


Accuracy: 0.35
```

Fig. 3. *Classification Performance of Kaggle Dataset*

The results presented in the classification report serve as additional proof of the model's efficiency. In most cases, there is a high precision rate along with a high recall rate for the model, which implies the accuracy and consistency of the model when making predictions. F1-scores are also quite balanced and show the model's reliability.

The best results are achieved in predicting Normal stress levels, while the performance remains stable for other types. The results obtained imply that the Random Forest approach is capable of recognizing the correlation between questionnaire answers and stress levels.

Confusion Matrix Analysis of Kaggle Dataset

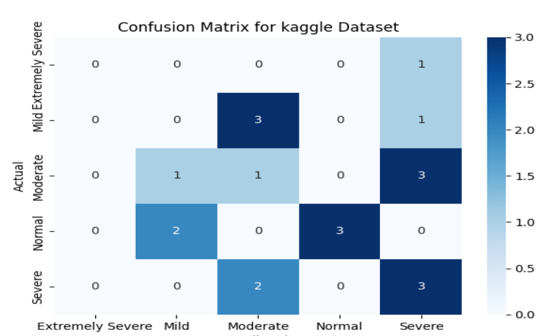

Fig. 4. *Confusion Matrix of Kaggle Dataset*

The confusion matrix derived from the Kaggle dataset shows that the classifier is able to predict stress levels in all five classes—Normal, Mild, Moderate, Severe, and Extremely Severe—in an efficient manner. The high prevalence of num- bers along the diagonal cells of the confusion matrix signifies that most data points have been accurately classified.

The classifier is able to predict stress categories very effi- ciently in case of more obvious stress categories like Normal and Severe. Some inaccuracies have been noticed by the model in predicting categories where the response pattern is similar; for example, Mild and Moderate.

Classification Performance of Real-Time Dataset

	precision	recall	f1-score	support
Extremely Severe	0.00	0.00	0.00	4
Mild	0.40	0.44	0.42	9
Moderate	0.15	0.18	0.17	11
Normal	0.90	0.67	0.77	27
Severe	0.31	0.67	0.42	6
accuracy			0.49	57
macro avg	0.35	0.39	0.35	57
weighted avg	0.55	0.49	0.51	57

Accuracy on Real-Time Dataset: 0.49122807017543857

Fig. 5. *Classification Performance of Real-time Dataset*

The classification report gives detailed information about the precision, recall, and F1-score of each class. From the report, the best performance was observed in the Normal class, with a precision of 0.90 and an F1-score of 0.77. This indicates that the model was reliable in classifying low-stress individuals. However, low performance was observed in the Moderate and Extremely Severe classes, with F1-scores of 0.17 and 0.00, respectively. This indicates that the model was unable to differentiate between high levels of stress, which could be due to class imbalance and similar patterns in features. The performance of the model was moderate in classifying the Mild and Severe classes, with F1-scores of 0.42.

Confusion Matrix Analysis of Real-Time Dataset

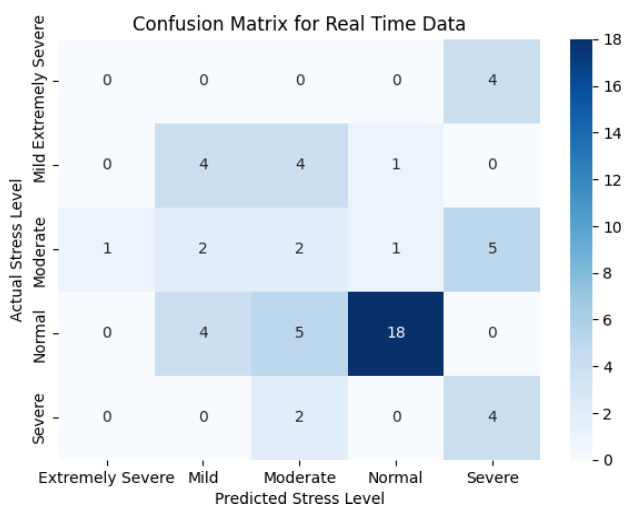

Fig. 6. *Confusion Matrix of Real-time Dataset*

The confusion matrix demonstrates how actual and pre- dicted stress levels compare in five categories: Normal, Mild, Moderate, Severe, and Extremely Severe. The model has performed well in predicting the Normal stress level, as there are 18 instances of accurate classification. However, misclassifications are noted in other categories, especially in Mild, Moderate, and Severe stress levels.

In addition, the model has failed to classify instances of Extremely Severe stress correctly, as there were no accurate predictions made. Furthermore, instances of Moderate stress were misclassified as Severe and Mild stress, indicating how complex it is to differentiate between stress levels.

Comparison of Kaggle and Real-Time Results

The comparison was made between the outcome produced by the Kaggle data and the real-time data set to determine the generalization ability of the model.

- Confusion Matrix Comparison: The confusion matrix in the Kaggle dataset is seen to have high diagonal dominance, reflecting many correct classifica- tions in all stress levels. On the other hand, the confusion matrix in the real-time dataset has a more dispersed pattern, implying more misclassification between neighboring stress levels like Mild, Moderate, and Severe. It can be noted that the machine learning algorithm functions better with well- organized data compared to real-time data.
- Classification Perfomance Comparison: The accuracy of classification using the Kaggle data set results in high values of precision, recall, and F1-scores for most categories. Conversely, the real-time data set features low values of all parameters, especially for Moderate and Extremely Severe stress level categories. The Normal category performs fairly well in both sets of data, implying that the classifier is highly accurate in recognizing unique stress levels. Intermediate categories experience a decline in accuracy owing to similarities in feature patterns.
- Accuracy Comparison: Accuracy attained from Kaggle dataset is much better compared to the real time dataset. The difference in the results shows that performance of the model heavily relies on the input data.

Discussion

The analysis demonstrates how although the model works efficiently with structured and labeled data, the efficiency drops once the model is applied to the real world because of factors such as noise, variance in responses, and any inconsistency that might arise in the case of self-reported surveys.

Conclusion

The current study proposed a machine learning-based ap- proach for predicting the levels of stress using the DASS-21 questionnaire. A random forest classifier was proposed and implemented for classifying the levels of stress severity into five classes: Normal, Mild, Moderate, Severe, and Extremely Severe. The proposed approach was trained and validated using an existing dataset and real-time survey data.

The proposed approach was able to achieve promising re- sults in identifying clear levels of stress severity, especially for the Normal category. However, the approach is still challenged in distinguishing between similar levels of stress severity, such as Mild and Moderate. The accuracy achieved for the proposed approach in real-time data indicates the feasibility of imple- menting machine learning-based approaches for automated stress assessment. This approach is considered a promising solution for the early identification of stress in individuals, especially in academic environments.

Future Scope

However, there are various ways in which the proposed system can be improved further. Firstly, the size and diversity of the dataset can improve the performance of the model significantly. In addition, various features such as behavioral data, lifestyle, and physiological data, including heart rate and sleep patterns, can provide more insight into the stress level of users.

Advanced machine learning and deep learning algorithms, such as neural network models, can also be explored for better classification accuracy. Moreover, the system can be designed as a real-time application or a mobile-based appli- cation for monitoring stress levels continuously. In addition, recommendation systems and mental health services can also be integrated for better assistance to users.

References

1. A. Shrivastava, M. Datta, A. DeSousa, O. Nayak, M. Gupta, S. Shrivas- tava, N. Shukla, H. Mange, J. Limbachia, S. Jagtap, M. Nemade and N. Shah, “Student Mental Health Screening: A Multidimensional, Digital Approach for Facilitating Early Intervention,” Journal of Psychiatry Research Reports, vol. 2, no. 2, pp. 1–8, 2025.
2. T. Chauhan and R. P. N, “Utilizing Machine Learning to Foster Em- ployee Mental Health in Modern Workplace Environment,” in Proc. IEEE Int. Students’ Conf. Electrical, Electronics and Computer Science (SCEECS), 2024.
3. A. Kapoor and S. Goel, “Predicting Stress at Workplace Using Machine Learning Techniques,” in Proc. 14th Int. Conf. Computing Communica- tion and Networking Technologies (ICCCNT), IEEE, 2023.
4. Y. Kumar et al., “Automated Classifications for Stress Symptoms Investigations and Stress Prediction Using Machine Learning Based Approaches,” in Proc. 2nd Int. Engineering Conf. Electrical, Energy, and Artificial Intelligence (EICEEAI), IEEE, 2023.
5. S. Verma, R. Kumar, P. Pani, A. Mohanty and P. Chaudhury, “Real Time Identification of Students’ Stress Factors Using Machine Learning Techniques,” in Proc. Int. Conf. Intelligent Computing and Sustainable Innovations in Technology (IC-SIT), IEEE, 2024.
6. S. Gjoreski, H. Gjoreski, M. Lusˇtrek and M. Gams, “Continuous Stress Detection Using a Wrist Device: In Laboratory and Real Life,” Proc. ACM Int. Joint Conf. Pervasive and Ubiquitous Computing, 2016.
7. J. Healey and R. Picard, “Detecting Stress During Real-World Driving Tasks Using Physiological Sensors,” IEEE Trans. Intelligent Transporta- tion Systems, vol. 6, no. 2, pp. 156–166, June 2005.