From Questionnaires to Actionable Insights: Machine Learning for Mental Stress Detection

Sumit Sudhakar Shinde
Department of Computer Engineering, Society for Computer Technology and Research, Pune Institute of Computer Technology, Savitribai Phule Pune University, Pune, Maharashtra, India
4440.sumitshinde@gmail.com (corresponding author)

Archana Santosh Ghotkar
Department of Computer Engineering, Society for Computer Technology and Research, Pune Institute of Computer Technology, Savitribai Phule Pune University, Pune, Maharashtra, India
aaghotkar@pict.edu

Received: 19 July 2025 | Revised: 30 July 2025, 28 August 2025, and 8 September 2025 | Accepted: 9 September 2025

 DOI: https://doi.org/10.48084/etasr.13513

ABSTRACT

Mental stress is a pervasive global health concern, necessitating timely and accurate detection for effective intervention and well-being. While questionnaire-based assessments are widely employed by medical practitioners, their efficacy can be influenced by questionnaire quality and assessor expertise. Addressing a notable research gap in the application of Machine Learning (ML) for mental stress assessment within the specific context of the Indian population, this study proposes a novel ML-based approach. Our methodology leverages comprehensive input data derived from Depression Anxiety and Stress-42 (DASS-42) questionnaire responses, Ten Item Personality Inventory (TIPI) questions, and relevant demographic factors. An ensemble voting classifier, integrating Logistic Regression (LR), Support Vector Machines (SVMs), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), was developed as the predictive model. Model robustness was rigorously evaluated using k-fold cross-validation, revealing consistent performance with a mean accuracy of 94.5% and a low standard deviation of 2.5%. Hyperparameters were meticulously tuned using grid search to optimize the ensemble's performance, resulting in a classification accuracy of 95% for mental stress detection. Furthermore, the model's predictions demonstrated a strong positive correlation (Pearson correlation coefficient of 0.822729) with results obtained from the standard Patient Health Questionnaire-9 (PHQ-9) questionnaire, statistically confirming its validity and alignment with established clinical assessments. This research offers a robust and validated decision support system that can aid mental health professionals in early diagnosis, guide customized preventive actions, and contribute significantly to destigmatizing mental health issues, thereby promoting overall mental well-being in diverse populations.

Keywords-mental stress; Machine Learning (ML); questionnaires; Depression Anxiety and Stress Scale-42 (DASS-42); Patient Health Questionnaire-9 (PHQ-9); ensemble classifiers

I. INTRODUCTION

Mental stress is a primary concern today, as it can severely affect an individual's physical and mental health. Many cultural and social factors also influence mental stress, including work-related pressures, discrimination, family disputes, and social challenges. To mitigate the negative impact of an unhealthy lifestyle and inadequate mental health, early detection of mental stress is critical for individual well-being and societal health.

Several studies have proposed the use of questionnaires to assess mental health, such as the Public Health Questionnaire-9 (PHQ-9) [1], Depression Anxiety and Stress Scale-42 (DASS-42) [2], Generalized Anxiety Disorder Scale-7 (GAD-7) [3], Alcohol Use Disorder (AUD) [4], and Major Depressive Disorder (MDD) [5].

This study proposes the use of Machine Learning (ML) algorithms to detect mental stress levels from the DASS-42 questionnaire, a widely used tool for measuring stress. We explore the effectiveness of several ML algorithms, including Logistic Regression (LR), Support Vector Machines (SVMs), Random Forest (RF), and Neural Networks (NNs), in predicting mental stress levels.

While ML has demonstrated considerable promise in healthcare domains, such as emotion recognition [6] and mental health prediction [7], there remains a significant research gap regarding its comprehensive application for mental stress detection by psychiatrists, particularly within the Indian population. Few studies have investigated ML-based mental health assessment in India. Authors in [8] reported a notable increase in anxiety and despair among Indian adolescents, with prevalence rates as high as 32% and they examined relationships with age, gender, geographic location, and school attendance. Understanding the impact of these factors on mental health is essential for developing effective interventions.

A. Social and Cultural Influences on Mental Stress

In addition to traditional questionnaires, considering social and cultural factors is essential [9], as these factors can significantly affect an individual's mental health. Table I summarizes key social and cultural influences and their potential effects on well-being.

TABLE I. IMPACT OF SOCIAL AND CULTURAL FACTORS ON MENTAL STRESS

Social/cultural factor	Impact on mental health
Social support	Lack of social support from society, friends, or family can increase stress levels, whereas supportive relationships can reduce stress.
Discrimination	Experiencing discrimination based on gender, race, sexuality, ethnicity, or socioeconomic status can lead to stressful events and negatively affect mental health.
Work and financial stress	Work-related issues, such as long hours, job insecurity, or financial pressure from loans and unemployment, contribute to stress.
Cultural norms	Cultural norms within specific religions or castes in particular regions can contribute to poor mental health.
Traumatic events	Exposure to traumatic events, such as family or workplace violence, can lead to long-term stress and negatively affect mental health.
Social status	Occupation, wealth, and education can limit access to resources and opportunities, increasing stress levels.
Social media	Excessive use or exposure to harmful online content can lead to anxiety and stress.

Considering these social and cultural factors is crucial for accurately assessing and addressing mental stress. Moreover, stigmatizing behavior remains a significant barrier to expressing emotions and seeking help, reducing an individual's willingness to access support or treatment for stress-related issues.

Traditionally, mental health has been assessed using various questionnaires. As ML demonstrates its effectiveness in building predictive models, researchers are increasingly developing automated solutions for mental health assessment, monitoring psychological activities, and analyzing human emotions, which serve as indicators of mental states.

II. RELATED WORK

Research on mental health in India has highlighted both the prevalence of stress and the potential for ML interventions. Authors in [10] investigated employee mental health in a specific Indian industry, evaluating its impact on productivity. Using questionnaires, they measured depression, anxiety, and stress among 90 workers. While no significant signs of depression were observed, 36% showed symptoms of anxiety, and 18% were affected by stress.

Indigenous communities face substantial health challenges, including mental health impacts, due to climate change. Authors in [11] analyzed 29 studies, showing that disruptions in environment, traditions, and way of life lead to grief, anxiety, and melancholy. This underscores the importance of considering local needs and adaptation efforts in policy and intervention planning.

ML and Artificial Intelligence (AI) have been widely applied to recognize emotions and predict mental health outcomes. Authors in [6] utilized datasets including facial images, videos, audio, and questionnaire responses, employing the DASS-21 questionnaire to identify anxiety and depression. They compared five classifiers, including SVM, Decision Trees, RF, Naïve Bayes, and k-Nearest Neighbors (KNN), demonstrating effective emotion recognition.

Authors in [7] investigated ML algorithms for predicting the severity of depression, anxiety, and stress using DASS-42. Comparing SVM and LR, they reported accuracies of 97.35%, 97.49%, and 97.20% for SVM and 98.15%, 98.05%, and 98.45% for LR, highlighting LR's superior performance.

The COVID-19 pandemic accelerated telemedicine adoption. Authors in [12, 13] proposed the DASS-CARE framework, integrating Internet of Things (IoT) and AI to enable decentralized, secure healthcare with real-time health monitoring, collaborative medical record access, and automated patient management.

Authors in [14] predicted depression severity and suicidal ideation using questionnaire and social media data. Their system classified severity into five stages, achieving 83.87% accuracy with Extreme Gradient Boosting (XGBoost) and 86.45% with LR.

Authors in [15] developed the BDI_Multi_Model, trained on the Beck Depression Inventory (BDI) questionnaire, achieving a Pearson correlation of 0.90 with official Canadian mental health survey data.

Authors in [16] developed the EKLS model for emotion recognition using facial expressions, achieving 99.82% accuracy through ensemble ML during the COVID-19 period.

Using a hybrid ConvNextBase and Light Gradient Boosting Machine (LightGBM) model, authors in [17] proposed a method for early Autism Spectrum Disorder (ASD) detection via eye-gaze analysis, achieving 95% accuracy. This study highlights the potential of combining deep learning feature extraction with gradient boosting for classification and underscores the utility of objective metrics in mental health diagnostics.

Authors in [18] explored attention-based eye-gaze analysis using transfer learning for autism prediction, demonstrating the potential of advanced deep learning models and eye-tracking features in non-clinical assessments. Although their focus was on autism, the study highlights the broader applicability of computer vision and transfer learning in mental health diagnostics.

Despite the growing body of ML research, studies on ML-based mental stress assessment by psychiatrists in India remain limited. Authors in [8, 10] suggest that ML can significantly improve mental health assessment, but existing studies often focus on narrow demographics or limited features, limiting the development of culturally relevant and robust diagnostic tools. The challenge is to design a consistent approach that addresses the limitations of traditional methods while providing meaningful insights.

To address this challenge, this study proposes an ML framework aimed at enhancing the reliability of mental stress detection. The framework also contributes to the development of an anti-stigma platform that empowers individuals to identify their stress levels.

The objectives of this research are:

- To develop and evaluate a robust ML-based model for mental stress detection using a dataset that includes DASS-42 questionnaire responses, Ten Item Personality Inventory (TIPI) questions, and relevant demographic factors from an Indian cohort.
- To assess the performance of an ensemble voting classifier (comprising LR, SVM, RF, and XGBoost) in accurately classifying mental stress levels, targeting high predictive accuracy and generalization capability.
- To validate the predictive model's output against an established clinical assessment tool, specifically the PHQ-9 questionnaire, demonstrating its potential as a reliable and complementary decision support system for mental health professionals in facilitating early diagnosis and guiding preventive interventions.

III. DEPRESSION ANXIETY AND STRESS SCALE-42 QUESTIONNAIRE

DASS-42 is a commonly used questionnaire to measure the severity of depression, anxiety, and stress symptoms. The DASS-42 is 4-point Likert scale questionnaire that consists of 42 items, assessing three domains: depression, anxiety, and stress. The items are rated on a scale from 0 to 3.

The DASS-42 questionnaire consists of three subscales:

1. Depression: The depression subscale assesses symptoms such as sadness, hopelessness, lack of interest, and low self-esteem. It consists of 14 items.
2. Anxiety: The anxiety subscale assesses nervousness, tension, and fear symptoms. It consists of 14 items.
3. Stress: The stress subscale assesses symptoms such as irritability, tension, and difficulty relaxing. It consists of 14 items.

The DASS-42 questionnaire has been extensively used in research and clinical settings and has demonstrated good reliability and validity. It can provide valuable insights for diagnosing and treating depression, anxiety, and stress-related disorders when used alongside other clinical information. However, it is not intended to replace a comprehensive clinical evaluation, and the results should be interpreted in conjunction with other clinical assessments.

To ensure the reliability of the DASS-42 in measuring depression, anxiety, and stress, its internal consistency is commonly evaluated using Cronbach's alpha reliability test.

A. *Cronbach's Alpha Reliability Test*

Cronbach's alpha measures the internal consistency reliability of a scale or questionnaire. It is widely used in social sciences, psychology, education, and other fields that utilize questionnaires or surveys to measure constructs or variables. The Cronbach's alpha coefficient ranges from 0 to 1, with higher values indicating greater internal consistency. An alpha of 0.70 or above is generally acceptable for most research purposes, although higher values are desirable. Cronbach' s alpha is calculated as follows:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum \text{variances of errors}}{\text{total variance}}\right) \quad (1)$$

where:

- n is the number of items in the scale or questionnaire.
- Calculate the mean and standard deviation of each item.
- Compute the correlation coefficients between each pair of items.
- Compute the total variance of the scale by summing the variances of all the items.
- The sum of variances of errors is calculated by subtracting the sum of the variances of each item from the total variance of the.

It is important to note that Cronbach's alpha assumes that the items measure the same construct and that the scale is unidimensional. For multidimensional scales, alternative reliability testing methods may be more appropriate.

IV. RESEARCH DESIGN AND METHODOLOGY

This study proposes an ML-based approach for detecting mental stress levels using questionnaire assessments. The approach utilizes a dataset of questionnaire responses from individuals experiencing mental stress and builds a predictive model using ML algorithms. Based on the responses, the predictive model is trained to classify the mental stress level. The ML algorithms employed in this study include LR, RF, KNN, SVM, XGBoost, and an ensemble voting classifier.

A. *System Architecture*

Figure 1 represents the system architecture for mental stress level detection using questionnaire data and ML. The process begins with the user interface, where participants provide responses to standardized questionnaires. These responses are transferred to the questionnaire processing module, which performs essential preprocessing steps such as data cleaning, feature extraction, and feature scaling to ensure data consistency and quality. The processed data are then used to train the ML algorithm, which classifies an individual's stress level. Supervised learning is used, as the dataset includes self-annotated labels representing predefined stress categories. The output of the ML algorithm is the stress level detected by the system. Finally, the system outputs the detected stress level.

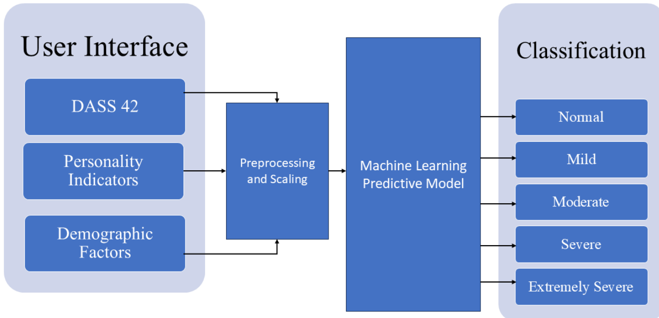

Fig. 1. Proposed system architecture for mental stress level detection using ML.

V. INTRODUCTION TO THE DATASET

The study primarily utilizes the 'DASS_data_21.02.19' dataset, publicly available on the Open Psychometrics website [19]. This dataset comprises survey responses collected between 2017 and 2019 via an online version of the DASS. The DASS questionnaire is formally described in [2].

The dataset includes responses from individuals who agreed to complete a research survey and indicated that their answers could be used for research purposes. The survey consists of items related to various aspects of mental health, including depression, anxiety, and stress. Participants were asked to rate the extent to which each item applied to them in the past week using a 4-point Likert scale.

In addition to the DASS items, the dataset also includes demographic information such as education, urban/rural background, gender, native language, age, hand preference, religion, sexual orientation, race, voting behavior, marital status, family size, and major (if applicable). Technical information, including the participant's country, screen size, unique network location, and source of the test, is also recorded.

The primary dataset used for model training consists of publicly available data, whereas the additional dataset used for validation was obtained through a survey conducted among voluntary participants. The survey responses, summarized in Table II, include data from 20 individuals belonging to an Indian cohort. These self-reported responses were used to validate the model's predictions and to perform correlation analysis between the predictive model outputs and the PHQ-9 questionnaire results.

TABLE II. SURVEY ANALYSIS ON QUESTIONNAIRE RESPONSES FROM THE INDIAN COHORT (20 PARTICIPANTS)

P_ID	Classification	Score	Stress level	Stress (%)	Anxiety level	Anxiety (%)	Depression level	Depression (%)
P1	N	5	N	60	N	40	N	0
P2	ML	33	ML	45.45454545	MD	33.33333333	N	21.21212121
P3	N	5	N	20	N	40	N	40
P4	N	19	N	26.31578947	ML	42.10526316	N	31.57894737
P5	S	80	S	33.75	ES	31.25	ES	35
P6	MD	61	MD	31.14754098	MD	21.31147541	ES	47.54098361
P7	ML	42	ML	35.71428571	MD	28.57142857	MD	35.71428571
P8	ES	91	S	36.26373626	ES	27.47252747	ES	36.26373626
P9	ML	32	N	40.625	N	15.625	MD	43.75
P10	MD	54	ML	33.33333333	ES	40.74074074	MD	25.92592593
P11	MD	55	MD	34.54545455	S	32.72727273	MD	32.72727273
P12	N	21	N	19.04761905	ML	38.0952381	N	42.85714286
P13	ML	52	ML	34.61538462	S	28.84615385	MD	36.53846154
P14	N	8	N	37.5	N	37.5	N	25
P15	N	8	N	37.5	N	25	N	37.5
P16	S	78	S	35.8974359	ES	30.76923077	S	33.33333333
P17	N	12	N	75	N	8.333333333	N	16.66666667
P18	S	66	S	43.93939394	S	27.27272727	MD	28.78787879
P19	ML	41	ML	39.02439024	MD	26.82926829	MD	34.14634146
P20	N	29	N	48.27586207	N	20.68965517	N	31.03448276

a. Footnote: N = Normal, ML = Mild, MD = Moderate, S = Severe, ES = Extremely Severe

A. Data Preprocessing

The dataset features used for the study include responses to various questions related to depression, anxiety, and stress, represented explicitly by items Q1A to Q42A on the DASS-42. These items explore aspects such as experiencing negative emotions, physical symptoms, difficulty in relaxation, and loss of interest or enjoyment. Additionally, the dataset includes responses to the TIPI, which provides insights into personality traits related to extraversion, criticality, dependability, anxiety, openness to new experiences, reservedness, sympathy, disorganization, emotional stability, and conventionality. Demographic information is also included in the dataset, including variables such as education level, urban/rural background, age, religion, race, marital status, family size, and major (if applicable).

These features offer a comprehensive perspective on the factors that may contribute to developing or exacerbating mental stress. By analyzing the relationships between depression levels and variables such as education, urban/rural background, age, religion, race, marital status, family size, and major, researchers can gain insights into the potential risk and contextual factors associated with stress symptoms.

The raw questionnaire responses and demographic information undergo several preprocessing steps to ensure data quality and suitability for ML algorithms:

- Missing value handling: Any missing data points within the features (DASS-42 items, TIPI, demographic factors) are addressed. For numerical features, imputation methods such as mean or median imputation are applied. For categorical features, mode imputation or a dedicated 'missing' category is used.
- Categorical feature encoding: Non-numerical demographic variables (e.g., gender, religion, marital status) are converted into a numerical format. One-hot encoding is applied to these categorical features to prevent the model from assuming an ordinal relationship where none exists.
- Feature scaling: To ensure that features with larger numerical ranges do not disproportionately influence the learning algorithms, MinMaxScaler is applied. This technique scales all features to a common range between 0 and 1.

VI. USE OF MACHINE LEARNING IN MENTAL STRESS DETECTION

ML plays a crucial role in assessing mental stress using various techniques and algorithms. It helps create models to predict individuals at risk of mental stress based on their demographic information, behavior patterns, and psychosocial factors. These models enable early intervention and support for those experiencing stress. ML techniques like feature selection and dimensionality reduction are used to identify the most critical factors contributing to mental stress assessment, making the models more accurate. Algorithms for recognizing emotions analyze facial expressions, voice patterns, and physiological signals to classify emotions associated with mental stress, allowing for real-time monitoring and timely interventions.

Additionally, ML models combined with Natural Language Processing (NLP) techniques can analyze text data from various sources, such as social media posts, chat transcripts, and clinical notes. This methodology helps assess mental stress levels, identifying language patterns that indicate distress or well-being. ML models provide a comprehensive understanding of mental stress by integrating data from multiple sources, such as physiological sensors, wearable devices, social media, and electronic health records. Personalized treatment recommendations are generated by considering individual characteristics, treatment history, and response patterns, tailoring interventions to meet specific needs.

ML models can also predict treatment outcomes and assess the risk of developing mental stress or related disorders by considering treatment type, duration, and patient characteristics. This technique aids in treatment planning, decision-making, and implementing preventive measures. Overall, ML empowers mental stress assessment using data-driven approaches and personalized interventions to support individuals in managing their mental well-being.

A. Ensemble Voting Classifier

An ensemble voting classifier combines multiple individual models and works by taking the predictions of various base models. This work considers LR, XGBoost, RF and SVM as the base models and combines them into a single prediction. There are different ways to combine the predictions, such as majority voting, weighted voting, or stacking. In this approach, majority voting is used.

Majority voting is the simplest and most common method, where the ensemble classifier selects the class that receives the most votes from the base models. For example, if there are three base models and two of them predict class A, whereas one predicts class B, the ensemble classifier will predict class A.

The ensemble voting classifier is a powerful technique that can improve the robustness and accuracy of classification models, especially in cases where individual models may have different strengths and weaknesses. However, choosing the right combination of base models and voting methods is essential to avoid overfitting by using cross-validation and regularization techniques.

1) Hyperparameter Tuning and Cross-Validation

To optimize the performance of the ensemble model and its base classifiers, grid search was employed for hyperparameter tuning. This systematic search technique explores a predefined range of hyperparameter values for each model. The optimization process was integrated with k-fold cross-validation (k = 5). In this procedure, the dataset is divided into five equal folds. The model is iteratively trained on four folds and validated on the remaining fold, rotating the validation fold in each iteration. The average performance across all five iterations provides a robust estimate of the model's generalization ability and guides the selection of optimal hyperparameters.

The specific hyperparameters tuned and selected for the base classifiers were:

- LR: Regularization parameter (C=10).
- XGBoost: Number of estimators (n_estimators=100), maximum tree depth max_depth=10.
- RF: Number of estimators (n_estimators=100), maximum tree depth (max_depth=10), minimum samples required to split an internal node (min_samples_split=2).
- SVM: Linear kernel, regularization parameter (C=10).

2) Statistical Analysis Rationale

The selection of statistical methods was driven by the need to rigorously validate the performance and reliability of our ML model:

- Descriptive statistics (accuracy, precision, recall, F1-score): These standard metrics were chosen to comprehensively evaluate the classification performance of our ensemble model and its base learners. They provide insights into the model's overall correctness, its ability to correctly identify positive instances, and its balance between these aspects.

- Correlation analysis (Pearson correlation coefficient): This method was used to quantify the linear relationship between the scores predicted by our ML model and the scores from the established PHQ-9 questionnaire. A strong correlation indicates alignment with a recognized clinical assessment tool.
- Hypothesis testing (ANOVA and U-test): These statistical tests were employed to formally assess whether the results from our ML model were statistically similar to the results obtained from the PHQ-9 questionnaire. The acceptance of the null hypothesis in these tests provides statistical evidence for the consistency of our model's predictions with a standard clinical procedure.

B. Proposed Algorithm for Mental Stress Detection

Algorithm 1 represents the pseudocode for developing an ML algorithm for predictive model building. Algorithm 2 describes the procedure used to calculate the severity level and percentage of anxiety, depression, and stress in an individual under assessment using the standard norms of the DASS-42 questionnaire [2]. Algorithm 3 illustrates the procedure for calculating the severity level of depression in an individual under assessment using the standard norms of the PHQ-9 questionnaire [1].

This work predicts the stress level of an individual, along with the severity level and percentage of anxiety, depression, and stress, as specified in Algorithm 2. The validation of the results predicted by the model, developed using the algorithmic procedure in Algorithm 1, is performed using the PHQ-9 results obtained from the algorithmic procedure in Algorithm 3. Algorithms 1 and 2 operate on the input from the DASS questionnaire responses, whereas Algorithm 3 operates on the PHQ-9 questionnaire responses. The results obtained from Algorithm 1 (proposed algorithm) are validated using Algorithm 3 for better generalization, as Algorithm 3 follows the standard procedure for assessing the primary mental health state of an individual.

```
Algorithm 1: Mental Stress Detection Predictive Model
Input: Dataset: A collection of labeled training and testing samples (e.g.,
       questionnaire responses with a 'target' mental stress level).
Output:P: The predicted class label (mental stress level) for a new user's responses.

# Phase 1: Train the Classifier Model
1. Split Data:
   // Extract the target mental stress level column and remove it from the dataset
   SET target TO Dataset['target']
   SET Dataset_features TO Dataset.drop('target', axis=1)
# Split dataset into training and testing sets
     (x_train, x_test, y_train, y_test) TO train_test_split(Dataset_features, target,
     test_size=0.2, random_state=42)
2. Scale Data:
// Initialize a Min-Max Scaler for feature scaling and scale training and testing data
   SET scaler TO MinMaxScaler()
   SET x_train_scaled TO scaler.fit_transform(x_train)
   SET x_test_scaled TO scaler.transform(x_test)
3. Train Ensemble Model:
   // Define base classifiers (e.g., LR, SVM, RF, XGBoost)
   SET Pred_model = VotingClassifier(estimators=[(model_1, ...), (model_n, ...)],
   voting='hard' or 'soft')
   Pred_model.fit(x_train_scaled, y_train)

# Phase 2: Accept Questionnaire Response from User and Predict
4. Prepare Questions and Answers:
   // Populate with possible questions and answers/options for each question
   SET list_of_questions TO [...]
   SET lists_of_answers TO [...]
5. Define Function to Get User Responses:
   DEFINE FUNCTION get_user_responses():
     SET user_answers_array = np.array([])
     FOR each_question, each_answer_options IN zip(list_of_questions, lists_of_answers):
       OUTPUT(each_question + '\n' + ' '.join(each_answer_options))
       SET user_input_str TO INPUT()              // Get user's answer as string input
       SET numerical_answer TO convert_to_numerical(user_input_str)
       user_answers_array.append(numerical_answer)
     RETURN user_answers_array
```

```
6. Collect User Responses:
   SET raw_user_responses TO get_user_responses()        // Call the function to collect
   responses
7. Scale User Responses:
   SET scaled_user_responses TO scaler.transform([raw_user_responses])
8. Predict Stress Level:
   SET P TO Pred_model.predict(scaled_user_responses)
9. Output Result:
   OUTPUT(P)

Algorithm 2: DASS-42 Questionnaire Assessment and Prediction Validation
Input: user_responses_DASS42: Set of numerical questionnaire responses (0-3 per item)
       for DASS-42, accepted from user.
Output:A1: Anxiety severity level, a_per: Anxiety percentage,
       D1: Depression severity level, d_per: Depression percentage,
       S1: Stress severity level, s_per: Stress percentage

# Prediction Validation using DASS-42 questionnaire assessment (based on standard DASS-
42 scoring)
1. Calculating the Anxiety Level & Percentage:
   SET A_indices TO [1, 3, 6, 8, 14, 18, 19, 22, 24, 27, 29, 35, 39, 40]  // Indices (0-
   indexed) of anxiety-related questions in DASS-42
   SET raw_anxiety_score = np.sum(user_responses_DASS42[A_indices])        // Sum scores
   for anxiety items
   SET max_anxiety_score_possible = 42        // Maximum possible raw score for anxiety
   subscale
   DEFINE FUNCTION get_anxiety_level(score):         // Maps raw score to severity level
     IF score <= 7 THEN RETURN "Normal"
     ELSE IF score <= 9 THEN RETURN "Mild"
     ELSE IF score <= 14 THEN RETURN "Moderate"
     ELSE IF score <= 19 THEN RETURN "Severe"
     ELSE RETURN "Extremely Severe"
   SET A1 TO get_anxiety_level(raw_anxiety_score)             // Determine anxiety level
   SET a_per TO (raw_anxiety_score / max_anxiety_score_possible) * 100     // Calculate
   anxiety percentage
   OUTPUT(A1, a_per)                          // Output the anxiety level and percentage
2. Calculating the Depression Level & Percentage:
   SET d_indices TO [2, 4, 9, 12, 15, 16, 20, 23, 25, 30, 33, 36, 37, 41]    // Indices
   (0-indexed) of depression-related questions in DASS-42
   SET raw_depression_score = np.sum(user_responses_DASS42[d_indices])    // Sum scores
   for depression items
   SET max_depression_score_possible = 42   // Maximum possible raw score for depression
   subscale
   DEFINE FUNCTION get_depression_level(score):      // Maps raw score to severity level
     IF score <= 9 THEN RETURN "Normal"
     ELSE IF score <= 13 THEN RETURN "Mild"
     ELSE IF score <= 20 THEN RETURN "Moderate"
     ELSE IF score <= 27 THEN RETURN "Severe"
     ELSE RETURN "Extremely Severe"
   SET D1 TO get_depression_level(raw_depression_score)    // Determine depression level
   SET d_per TO (raw_depression_score / max_depression_score_possible) * 100   //
   Calculate depression percentage
   OUTPUT (D1, d_per)                      // Output the depression level and percentage
3. Calculating the Stress Level & Percentage:
   SET s_indices TO [0, 5, 7, 10, 11, 13, 17, 21, 26, 28, 31, 32, 34, 38]  // Indices of
   stress-related questions in DASS-42
   SET raw_stress_score = np.sum(user_responses_DASS42[s_indices])     // Sum scores for
stress items
```

```
   SET max_stress_score_possible = 42  // Maximum possible raw score for stress subscale
   DEFINE FUNCTION get_stress_level(score):          // Maps raw score to severity level
     IF score <= 14 THEN RETURN "Normal"
     ELSE IF score <= 18 THEN "Mild"
     ELSE IF score <= 25 THEN RETURN "Moderate"
     ELSE IF score <= 33 THEN RETURN "Severe"
     ELSE RETURN "Extremely Severe"
   SET S1 TO get_stress_level(raw_stress_score)                // Determine stress level
   SET s_per TO (raw_stress_score / max_stress_score_possible) * 100 // Calculate stress
   percentage
   OUTPUT(S1, s_per)                               // Output the stress level and percentage

Algorithm 3: PHQ-9 Questionnaire Prediction Validation
Input: user_responses_PHQ-9: Set of numerical PHQ-9 questionnaire responses (0-3 per
item), accepted from user.
Output:P1: The class label (depression severity level) for accepted responses

# Prediction Validation with PHQ-9 questionnaire (based on standard PHQ-9 scoring)
1. Prepare Questions and Answers (assumed to be pre-defined or loaded):
   SET list_of_questions TO [...]                  // Populate with the 9 PHQ-9 questions
   SET lists_of_answers TO [...]      // Populate with possible answers/options for each
   question
2. Define Function to Get User Responses:
   DEFINE FUNCTION get_PHQ-9_responses():
     SET PHQ-9_answers_array = np.array([])    // empty NumPy array for user's numerical
     answers
     FOR each_question, each_answer_options IN zip(list_of_questions, lists_of_answers):
       OUTPUT(each_question + '\n' + ' '.join(each_answer_options)) // Display question
       and its options
       SET user_input_str TO INPUT()                       // Get user's answer as string
       input
       SET numerical_answer TO convert_to_numerical(user_input_str)                //
       Convert user_input_str to numerical value based on PHQ-9 scoring (0-3)
       PHQ-9_answers_array.append(numerical_answer)       // Append numerical answer to
       the array
     RETURN PHQ-9_answers_array       // Return the array of user's numerical responses
3. Define Condition Function:
   DEFINE FUNCTION get_PHQ-9_level(score):         // Maps raw PHQ-9 score to depression
   severity level
     IF score <= 4 THEN RETURN "Minimal Depression"
     ELSE IF score <= 9 THEN RETURN "Mild Depression"
     ELSE IF score <= 14 THEN RETURN "Moderate Depression"
     ELSE IF score <= 19 THEN RETURN "Moderately Severe Depression"
     ELSE RETURN "Severe Depression"
4. Process Responses and Predict:
   SET raw_PHQ-9_responses TO get_PHQ-9_responses()  // Collect user responses for PHQ-9
   SET total_PHQ-9_score TO np.sum(raw_PHQ-9_responses)     // Calculate the sum of PHQ-
   9 scores
   SET P1 TO get_PHQ-9_level(total_PHQ-9_score)
5. Output Result:
   OUTPUT(P1)                                         // Print the predicted depression level
```

VII. RESULTS AND DISCUSSION

The proposed approach was evaluated using questionnaire responses collected from 20 voluntary participants at an engineering college in India. Predictive classification was performed using the developed ML model, and the results are summarized in Table III. Among all models, the ensemble voting classifier achieved the highest accuracy of 95%. The stress levels predicted by the model showed a strong correlation with the PHQ-9 questionnaire results for the same participants, with a Pearson correlation coefficient of 0.822729, as presented in Table IV.

TABLE III. PERFORMANCE EVALUATION OF ML PREDICTIVE MODELS

Classifier	LR	RF	KNN	SVM	XGBoost	Voting classifier
Precision	0.89	0.78	0.75	0.91	0.85	0.95
Recall	0.89	0.77	0.76	0.90	0.85	0.95
F1-score	0.89	0.77	0.75	0.90	0.85	0.95
Accuracy	0.89	0.78	0.76	0.91	0.85	0.95

TABLE IV. CORRELATION ANALYSIS BETWEEN THE PREDICTIVE MODEL SCORE AND THE PHQ-9 SCORE ON SURVEY RESPONSES

P_ID	Predictive model score	PHQ-9 score
1	5	4
2	33	12
3	5	4
4	19	4
5	80	13
6	61	20
7	42	8
8	91	24
9	32	8
10	54	10
11	55	11
12	21	9
13	52	8
14	8	4
15	8	2
16	78	10
17	12	4
18	66	12
19	41	10
20	29	8
Pearson correlation coefficient	0.822729	

The experimental results indicate that the voting classifier provides the best performance in detecting mental stress. Figure 2 presents the precision–recall curve for this classifier. The individual classifiers, LR, RF, KNN, SVM, and XGBoost, also performed reasonably well, with accuracies of 89%, 78%, 75%, 91%, and 85%, respectively, as shown in Table III. Additionally, a comparison with traditional statistical methods, specifically ANOVA and U-test, demonstrates that ML models outperform conventional approaches in mental stress detection (Table V). The robustness of the model was further validated using k-fold cross-validation. The ensemble classifier exhibited consistent performance across folds, achieving a mean accuracy of 0.945 and a standard deviation of 0.025, indicating minimal variation across subsets. The hyperparameters were optimized using grid search in combination with 5-fold cross-validation.

TABLE V. HYPOTHESIS TESTING FOR VALIDATION OF THE PREDICTIVE MODEL ON SURVEY QUESTIONNAIRE RESPONSES

Test	P (calculated)	P (tabulated)	Hypothesis test result
ANOVA	0.0450	4.098	H_0 accepted
U-test	186.5	127	H_0 accepted

a. Null hypothesis (H0): The ML predictive model results are approximately similar to PHQ-9 results.

b. Alternate hypothesis (Ha): The ML predictive model results are not similar to PHQ-9 results.

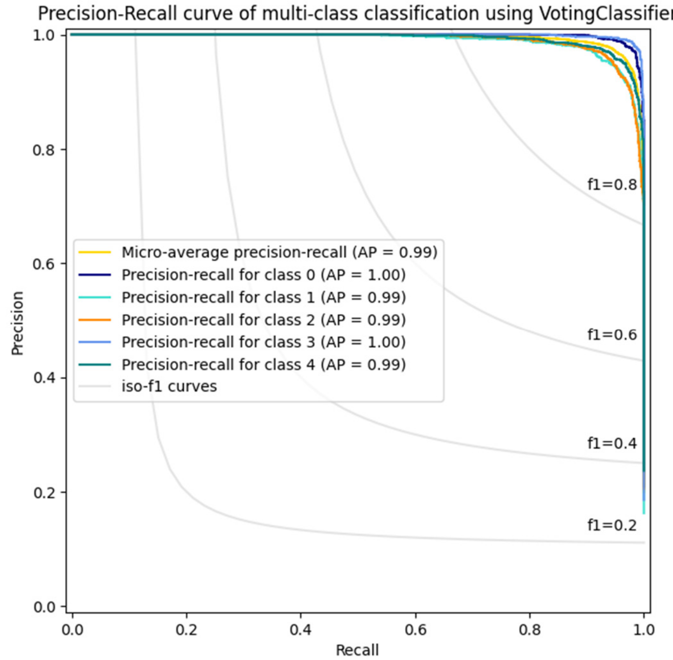

Fig. 2. Precision–recall curve for multi-class classification using the ensemble voting classifier.

The ensemble model utilized four base classifiers: LR (C=10), XGBoost (n_estimators=100, max_depth=10), RF (n_estimators=100, max_depth=10, min_samples_split=2), and SVM with a linear kernel (C=10). The larger C values in LR and SVM enhanced model flexibility. By combining DASS-42 questionnaire responses with personality indicators and demographic features, the proposed approach improves mental stress detection accuracy and provides decision support for mental health professionals. Further improvement may include additional features, such as heart rate variability, facial expressions, and speech patterns, for multimodal assessment.

A novel aspect of this study is the application of the ensemble methodology to an Indian cohort, which provides culturally specific and contextually relevant insights into patterns of mental stress. This approach addresses a notable research gap by exploring machine learning (ML)-based stress assessment within a demographic that has been underrepresented in prior studies

A. Comparative Study

Table VI presents a comparative analysis between the proposed methodology and several existing approaches for mental stress detection. The proposed method integrates the DASS-42 questionnaire, personality indicators, and demographic factors within an ensemble voting classifier framework. The ensemble voting classifier achieved a notably high accuracy of 95%, outperforming traditional approaches.

Furthermore, the model's robustness and generalizability were validated through k-fold cross-validation, yielding a mean accuracy of 0.945 with a standard deviation of 0.025, confirming the reliability and effectiveness of the proposed predictive system for real-world mental stress assessment.

TABLE VI. COMPARISON OF PROPOSED AND EXISTING METHODOLOGIES

Ref.	Dataset & Source	Methodology (Classifier)	Accuracy (%)
Proposed method	DASS-42, TIPI, Demographics (Open Psychometrics)	Ensemble voting classifier	95
[7]	University student depression data	Bootstrapping technique with Relief feature selection	93.16
[14]	Real-time data from students and parents (questionnaires similar to PHQ-9)	XGBoost, LR	86.45
[20]	Socio-demographic and occupational information	CatBoost	89.30
[21]	Socio-demographic and psychosocial information	AdaBoost	92.56
[22]	Occupational and socio-demographic information	RF	85.5

VIII. LIMITATIONS

While this study presents a promising ML approach for mental stress detection, several limitations should be considered when interpreting the findings and assessing their generalizability:

- Dataset characteristics: The study relied on a publicly available dataset composed primarily of self-reported questionnaire responses. Such self-report data may be subject to biases, including social desirability, recall errors, or misinterpretation of questions, potentially affecting the accuracy of the ground truth labels.
- Specific cohort: The survey data were collected from voluntary participants, primarily from an engineering college in India. While this provides valuable insights into this specific demographic, the findings may not be directly generalizable to broader populations with different socio-cultural backgrounds, age groups, or professional contexts without further validation.
- Preliminary nature of findings: Although the model demonstrates high accuracy on the utilized dataset, the analysis was based on a limited sample of responses. While the full training dataset was substantially larger, external validation using independent and diverse datasets is essential to confirm the model's robustness and generalizability across varied real-world scenarios.
- Feature scope: Incorporating additional physiological signals or behavioral data could provide a more holistic understanding of mental stress and further improve predictive performance.

IX. CONCLUSION

Mental stress is a significant global health challenge, and traditional assessment methods often face limitations due to their subjective nature and reliance on expert interpretation. This study addressed a notable knowledge gap by proposing a Machine Learning (ML)-based framework for assessing mental stress levels within the context of the Indian population, leveraging comprehensive data from the Depression Anxiety and Stress-42 (DASS-42) questionnaire, the Ten Item Personality Inventory (TIPI), and various demographic factors.

This framework successfully integrates and preprocesses these data. Results demonstrate that the ensemble voting classifier ML model effectively classifies mental stress levels with an accuracy of 95%. The key findings of this study confirm the feasibility of using ML for objective mental health assessment. The novelty of this approach lies in the combination of these specific questionnaires with demographic data from a previously underrepresented population, providing a new data-driven perspective on mental stress analysis.

The primary contributions of this work are twofold:

1. The development of a robust data preprocessing pipeline tailored for questionnaire and demographic data.
2. The validation of ML models for accurately identifying different levels of mental stress, thereby offering a practical and scalable tool for early detection and intervention.

This work marks a significant step toward creating a more accessible and objective system for mental health screening.

CONFLICT OF INTEREST

The authors have no relevant financial or non-financial interests to disclose.

ETHICAL CONSIDERATIONS

This study utilized the publicly available DASS_data_21.02.19 dataset for training the ML model. In addition, all individuals participating in the surveys provided informed consent, ensuring that they understood the purpose of the study and that their privacy was protected. As this study employs a publicly available dataset and obtained explicit consent from all survey participants, ethical approval from an institutional review board or ethics committee was not required.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are publicly available from the Open Psychometrics Project (https://openpsychometrics.org/_rawdata/) [19]. The DASS questionnaire, which forms the basis of the dataset, is formally described in [2].

REFERENCES

[1] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9," *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, Sep. 2001, https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

[2] S. H. Lovibond and P. F. Lovibond, *Manual for the Depression Anxiety Stress Scales*, 2nd ed. Sydney, Australia: Psychology Foundation of Australia, 1995.

[3] R. L. Spitzer, K. Kroenke, J. B. W. Williams, and B. Löwe, "A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7,"

Archives of Internal Medicine, vol. 166, no. 10, pp. 1092–1097, May 2006, https://doi.org/10.1001/archinte.166.10.1092.

[4] P. Anderson *et al.*, "Improving the delivery of brief interventions for heavy drinking in primary health care: outcome results of the Optimizing Delivery of Health Care Intervention (ODHIN) five-country cluster randomized factorial trial," *Addiction*, vol. 111, no. 11, pp. 1935–1945, Nov. 2016, https://doi.org/10.1111/add.13476.

[5] A. J. Mitchell and J. C. Coyne, "Do ultra-short screening instruments accurately detect depression in primary care?: A pooled analysis and meta-analysis of 22 studies," *British Journal of General Practice*, vol. 57, no. 535, pp. 144–151, Feb. 2007.

[6] A. Singh and D. Kumar, "Identification of Anxiety and Depression Using DASS-21 Questionnaire and Machine Learning," in *2021 First International Conference on Advances in Computing and Future Communication Technologies*, Meerut, India, 2021, pp. 69–74, https://doi.org/10.1109/ICACFCT53978.2021.9837365.

[7] K. S. Srinath, K. Kiran, S. Pranavi, M. Amrutha, P. D. Shenoy, and K. R. Venugopal, "Prediction of Depression, Anxiety and Stress Levels Using Dass-42," in *2022 IEEE 7th International conference for Convergence in Technology*, Mumbai, India, 2022, pp. 1–6, https://doi.org/10.1109/I2CT54291.2022.9824087.

[8] K. Singh, M. Junnarkar, and S. Sharma, "Anxiety, stress, depression, and psychosocial functioning of Indian adolescents," *Indian Journal of Psychiatry*, vol. 57, no. 4, Dec. 2015, Art. no. 367, https://doi.org/10.4103/0019-5545.171841.

[9] S. Amudhan and G. Gopalkrishna, "Depression in India: Let's talk." World Health Organization, 2017.

[10] S. Rao and N. Ramesh, "Depression, anxiety and stress levels in industrial workers: A pilot study in Bangalore, India," *Industrial Psychiatry Journal*, vol. 24, no. 1, Jun. 2015, Art. no. 23, https://doi.org/10.4103/0972-6748.160927.

[11] M. Sahu, B. Chattopadhyay, R. Das, and S. Chaturvedi, "Measuring Impact of Climate Change on Indigenous Health in the Background of Multiple Disadvantages: A Scoping Review for Equitable Public Health Policy Formulation," *Journal of Prevention*, vol. 44, no. 4, pp. 421–456, Aug. 2023, https://doi.org/10.1007/s10935-022-00718-8.

[12] J. N. Al-Karaki, A. Gawanmeh, M. Ayache, and A. Mashaleh, "DASS-CARE: A Decentralized, Accessible, Scalable, and Secure Healthcare Framework using Blockchain," in *2019 15th International Wireless Communications & Mobile Computing Conference*, Tangier, Morocco, 2019, pp. 330–335, https://doi.org/10.1109/IWCMC.2019.8766714.

[13] M. Ayache, A. Gawanmeh, and J. N. Al-Karaki, "DASS-CARE 2.0: Blockchain-Based Healthcare Framework for Collaborative Diagnosis in CIoMT Ecosystem," in *2022 5th Conference on Cloud and Internet of Things*, Marrakech, Morocco, 2022, pp. 40–47, https://doi.org/10.1109/CIoT53061.2022.9766532.

[14] S. Jain, S. P. Narayan, R. K. Dewang, U. Bhartiya, N. Meena, and V. Kumar, "A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter," in *2019 IEEE Students Conference on Engineering and Systems*, Allahabad, India, 2019, pp. 1–6, https://doi.org/10.1109/SCES46477.2019.8977211.

[15] R. S. Skaik and D. Inkpen, "Predicting Depression in Canada by Automatic Filling of Beck's Depression Inventory Questionnaire," *IEEE Access*, vol. 10, pp. 102033–102047, 2022, https://doi.org/10.1109/ACCESS.2022.3208470.

[16] B. Sahoo and A. Gupta, "An Ensemble Kernelized-based Approach for Precise Emotion Recognition in Depressed People," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18873–18882, Dec. 2024, https://doi.org/10.48084/etasr.8785.

[17] R. Bidwe, S. Mishra, S. Bajaj, and K. Kotecha, "Leveraging hybrid model of ConvNextBase and LightGBM for early ASD detection via eye-gaze analysis," *MethodsX*, vol. 14, Jun. 2025, Art. no. 103166, https://doi.org/10.1016/j.mex.2025.103166.

[18] R. Vasant Bidwe, S. Mishra, S. Kamini Bajaj, and K. Kotecha, "Attention-Focused Eye Gaze Analysis to Predict Autistic Traits Using Transfer Learning," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, May 2024, Art. no. 120, https://doi.org/10.1007/s44196-024-00491-y.

[19] "Open psychology data: Raw data from online personality tests." Openpsychometrics. https://openpsychometrics.org/_rawdata/.

[20] A. Sau and I. Bhakta, "Screening of anxiety and depression among seafarers using machine learning technology," *Informatics in Medicine Unlocked*, vol. 16, Jan. 2019, Art. no. 100228, https://doi.org/10.1016/j.imu.2019.100228.

[21] Md. S. Zulfiker, N. Kabir, A. A. Biswas, T. Nazneen, and M. S. Uddin, "An in-depth analysis of machine learning approaches to predict depression," *Current Research in Behavioral Sciences*, vol. 2, Nov. 2021, Art. no. 100044, https://doi.org/10.1016/j.crbeha.2021.100044.

[22] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," *Procedia Computer Science*, vol. 167, pp. 1258–1267, Jan. 2020, https://doi.org/10.1016/j.procs.2020.03.442.