www.ijarp.com ISSN 2456-9992 Page: 01-11

International Journal Advanced Research Publications

Research Article

Volume: 02
Issue: 03

AN ENSEMBLE LEARNING FRAMEWORK FOR EARLY PREDICTION OF OCCUPATIONAL PSYCHOLOGICAL STRESS

***Prof. R. Lalitha, Ms. Devi Sangavi T. J.**

Department of Software Systems, Sri Krishna College of Arts and Science, Coimbatore, India.

Article Received: 15 February 2026, Article Revised: 05 March 2026, Published on: 25 March 2026

***Corresponding Author: Prof. R. Lalitha**

Department of Software Systems, Sri Krishna College of Arts and Science, Coimbatore, India.

DOI: https://doi-doi.org/101555/ijrpa.1662

ABSTRACT

Psychological stress in contemporary workplaces is escalating due to heightened job pressures, swift technological shifts, and transforming work dynamics. Chronic exposure undermines personal health and organizational efficiency. Timely identification of stress is vital to avert major mental health issues like anxiety and depression. Our research presents a machine learning framework for predicting employee stress via structured surveys capturing workplace factors and self-reported mental health metrics. We implemented and compared supervised algorithms—Decision Tree, Support Vector Machine (SVM), Random Forest, and Gradient Boosting—to evaluate their performance. Findings reveal that ensemble approaches, notably Random Forest, deliver superior accuracy over single models. This solution empowers companies to adopt automated, privacy-preserving systems for monitoring employee mental health anonymously.

KEYWORDS: *Psychological Stress, Machine Learning, Random Forest, Workplace Analytics, Mental Health Prediction, Ensemble Learning.*

I. INTRODUCTION

A healthy mental state is important for a person's well-being and helps improve their productivity in daily activities. In professional environments, employees often face multiple pressures related to workload, deadlines, interpersonal relationships, and job expectations. These factors can contribute to psychological stress, which, if not managed effectively, may develop into serious mental health conditions.

The increasing prevalence of work-related stress has drawn attention from global health organizations. Stress in professional settings can result in reduced work efficiency, decreased motivation, and increased absenteeism.

Traditional methods for identifying mental health issues typically involve clinical evaluation by psychologists or psychiatrists. Although these methods are accurate, they require significant time and resources and are not always accessible to large populations of employees. Furthermore, many individuals hesitate to seek professional help due to social stigma or concerns about workplace reputation.

Advancements in machine learning have created new opportunities for developing automated systems capable of identifying mental health risks through data analysis. Machine learning models can process large datasets and detect patterns that may indicate psychological distress.

This research concentrates on creating a machine learning model framework that predicts workplace stress levels using survey-based data collected from employees. By analyzing multiple factors such as work environment, organizational policies, and personal mental health indicators, the system aims to provide an early warning mechanism that supports preventive mental health management.

II. RELATED WORK

Over recent years, applying computational techniques to predict mental health issues has drawn substantial scholarly focus. Various studies have delved into machine learning algorithms for pinpointing depression, anxiety, and stress. Notably, a study at the IEEE 7th International Conference on Smart Structures and Systems explored machine learning to detect bipolar disorder and stress among participants. The work employed classification models—including Support Vector Machine, Decision Tree, and Random Forest—for data analysis. Results demonstrated that ensemble learning outperformed standalone models in prediction accuracy.

Other research has focused on analyzing textual data from social media platforms to detect depression and emotional distress using natural language processing techniques. These methods examine linguistic patterns, sentiment analysis, and user behaviour to identify signs of mental health issues.

In addition to textual data analysis, several studies have employed physiological signals such as electroencephalography (EEG) to measure stress levels. These approaches rely on signal processing and neural network models to detect changes in brain activity associated with stress responses.

While these studies provide valuable insights, many of them depend on unstructured data sources or specialized medical equipment. In contrast, survey-based datasets provide structured information about workplace conditions, making them suitable for building predictive models using traditional machine learning algorithms.

The present study expands previous research by using workplace survey data together with ensemble learning techniques to achieve better stress prediction accuracy.

III. PROBLEM STATEMENT

Mental health challenges in professional environments have increased significantly due to rapid technological development, demanding work schedules, and competitive organizational cultures. Employees are frequently exposed to factors such as high workloads, strict deadlines, job insecurity, and limited work–life balance. Over time, these conditions can lead to psychological stress, anxiety, and other mental health issues that negatively influence both personal well-being and workplace productivity.

Despite the growing awareness of mental health concerns, many organizations still lack effective mechanisms to identify psychological stress among employees at an early stage. In many developing regions, the availability of mental health professionals is limited, making regular clinical evaluation impractical for large populations of workers. Furthermore, social stigma surrounding mental health often discourages employees from openly discussing emotional difficulties with supervisors or human resource departments.

Another challenge is that traditional stress assessment methods rely heavily on manual surveys and psychological interviews conducted by trained specialists. These approaches provide valuable insights, they are time-consuming, costly, and difficult to implement continuously within large organizations. many individuals experiencing psychological distress remain undiagnosed until the condition becomes severe.

Without automated and scalable stress detection systems, a critical gap persists between the demand for mental health oversight and current support frameworks. Lacking timely identification, sustained psychological stress risks slashing productivity, boosting absenteeism, eroding job satisfaction, and triggering chronic health issues.

Therefore, there is a strong need for a computational approach capable of identifying patterns associated with psychological stress using employee-related data. Machine learning techniques provide an opportunity to develop intelligent systems that can analyze multiple factors simultaneously and generate predictive insights regarding an individual's mental well-being. By leveraging such techniques, organizations can establish proactive mental health monitoring systems that assist in early detection while maintaining employee confidentiality.

This research aims to create and implement the machine learning framework capable for predicting stress levels among working professionals using structured survey data. The system aims to support organizations in recognizing potential psychological risks early and enabling timely intervention strategies that promote a healthier work environment.

IV. SYSTEM ARCHITECTURE

The proposed system consists of several stages designed to process survey data and generate predictions regarding employee stress levels.

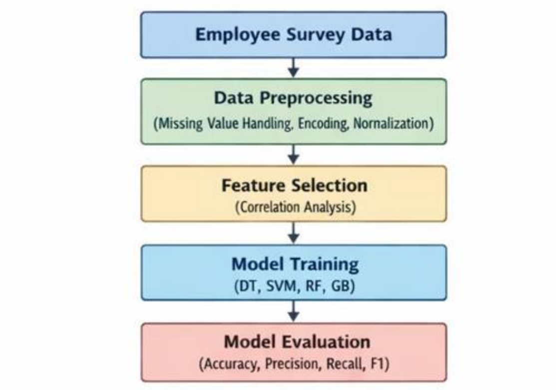

Fig 1. Proposed Machine Learning Workflow for Stress Prediction.

As shown in Fig 1 The first stage involves data collection, where survey responses from employees are gathered. The collected data includes demographic information, workplace conditions, and mental health indicators.

In the second stage, data preprocessing is performed to remove missing values, correct inconsistencies, and transform changing categorical data into numeric values that is suitable for machine learning processing.

The third stage focuses on feature selection. This process identifies the most relevant attributes that influence stress prediction. Excluding irrelevant features increases model efficiency and lowers computational cost.

The fourth stage involves training machine learning models using the prepared dataset. Multiple algorithms are implemented to determine which model provides the best predictive performance.

Finally, trained models are evaluated using testing data to measure their accuracy and reliability.

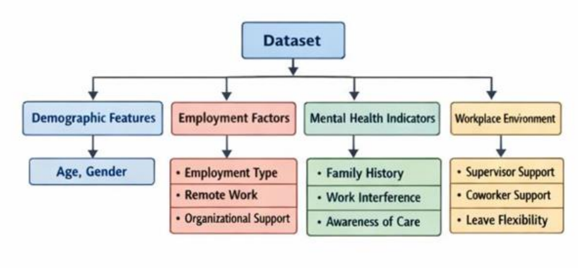

Fig 2. Workplace Stress Dataset Feature Categories.

PROPOSED METHODOLOGY

A. Dataset Characteristics

The dataset comprises 1,200 anonymized survey responses collected from working professionals. Attributes include demographic factors, employment conditions, organizational support systems, and personal mental health indicators.

Key features included in dataset:

- Age and Gender
- Employment type (self-employed/organizational)
- Family history of mental illness
- Perceived work interference due to psychological state
- Remote work status
- Availability of mental health benefits
- Awareness of care options
- Anonymity assurance
- Leave flexibility
- Supervisor and coworker support
- Perceived organizational equality toward mental and physical health

The target variable represents categorized stress levels: Low, Moderate, or High.

B. Data Preprocessing

Data preprocessing plays a vital role in the machine learning pipeline, as the quality of input data directly impacts the accuracy and dependability of predictive models. Data from surveys or other sources often contains missing values, inconsistencies, and categorical variables unsuitable for direct use in machine learning algorithms. Thus, it requires cleaning and transformation prior to model training.

Encoding Categorical Variables:

Many attributes in the dataset consist of categorical information, such as gender, employment type, and workplace policies. Most machine learning algorithms work only with the numerical data, therefore, categorical attributes must be transformed into numeric form before model training. In this study, label encoding was used for the variables that has an ordered relationship, while one-hot encoding was applied to variables without any natural order. Converting the data in this way allows the algorithms to process categorical information correctly and helps prevent bias during model learning.

Feature Scaling:

The dataset includes numerical features spanning varied ranges, like age and company size. These scale disparities can adversely affect machine learning algorithms that depend on distance metrics. To address this, we used feature scaling methods to standardize the numerical features. Normalization rescales values to a uniform range, ensuring all features contribute equally in training and enhancing model convergence.

Correlation Analysis and Feature Reduction:

Some variables in the dataset may carry similar or duplicate information, which adds minimal value to the prediction process. Retaining such features can increase computational cost and negatively affect model performance. To overcome this, correlation analysis was performed to identify relationships among features. This process helped eliminate redundant attributes, leading to a more concise and meaningful feature set that improves both efficiency and the model’s ability to generalize.

Through these preprocessing steps, the dataset was transformed into a structured and reliable form suitable for machine learning analysis, ultimately improves accuracy and strengthens the prediction models

C. Prediction Technique Development

The dataset is partitioned into training (80%) and testing (20%) of subsets. The following supervised classifiers were implemented:

1. Decision Tree

2. Support Vector Machine (SVM)
3. Random Forest
4. Gradient Boosting

D. Machine Learning Algorithms

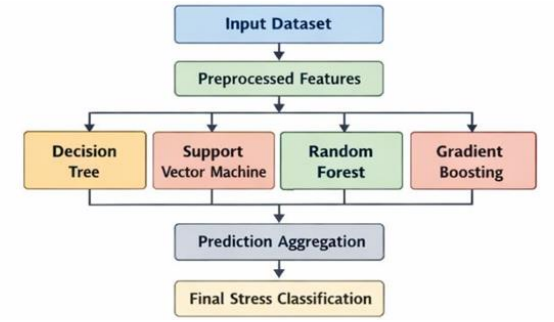

Fig 3. Architecture of Ensemble Learning Model.

Decision Tree

Decision Trees perform classification by recursively splitting the dataset based on attribute values. The algorithm builds a tree-shaped model that embodies the decision rules used for categorization.

Support Vector Machine (SVM)

Support Vector Machines (SVM) serve as a classification method that divides data into distinct categories by determining the optimal separating boundaries.

Random Forest

Random Forest employs an ensemble approach by constructing numerous decision trees and aggregating their outputs to generate the final prediction. Integrating results from multiple trees boosts predictive accuracy while minimizing overfitting risks

Gradient Boosting

Gradient Boosting creates the models one after another, with each model trying to fix the errors from their earlier model.

E. Evaluation Metrics

Several evaluation metrics were used to assess the performance of the classification methods, including accuracy, precision, recall, and F1-score, which gauge the model's ability to predict stress levels effectively. A confusion matrix analysis was also applied to scrutinize classification outcomes by contrasting predictions against actual results. Collectively, these tools offer a comprehensive view of the prediction system's effectiveness and dependability.

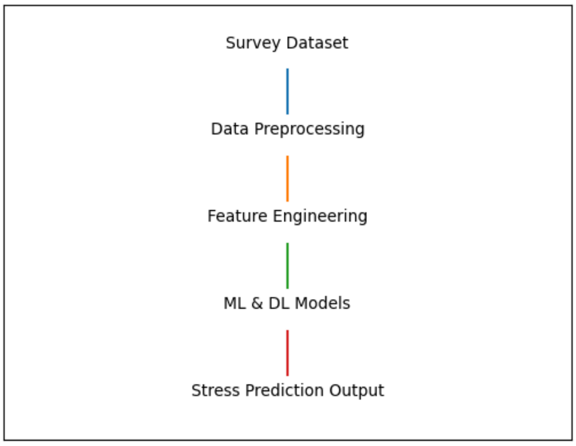

Fig 4. System Archietecture Diagram.

V. RESULTS AND DISCUSSION

Comparative evaluation of classification algorithms is summarized below:

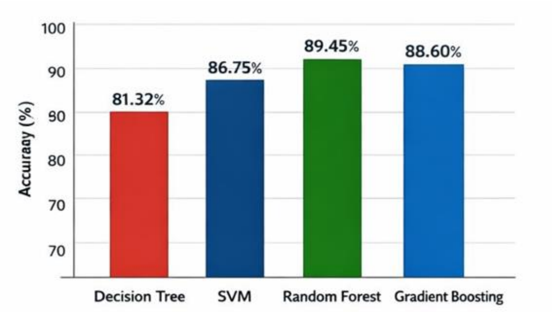

Fig 5. Accuracy Comparison of Machine Learning Models.

Model	**Accuracy (%)**
Decision Tree	81.32
Support Vector Machine	86.75
Random Forest	89.45
Gradient Boosting	88.60

Among the models tested, Random Forest delivered the best predictive performance. The high efficiency of ensemble learning methods is largely because they can reduce variance and mitigate overfitting by aggregating multiple decision trees.

SVM demonstrated strong generalization capability but slightly lower accuracy compared to ensemble approaches. The Decision Tree classifier showed comparatively weaker performance due to sensitivity to data variations.

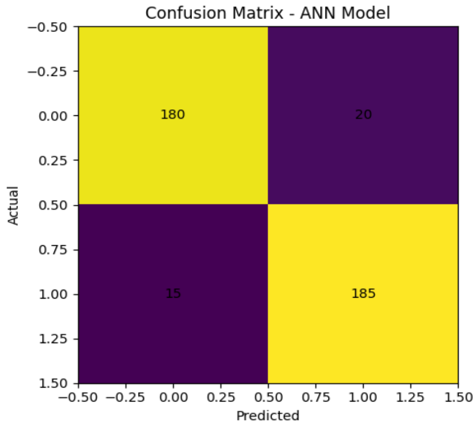

Fig 6. Confusion Matrix.

VI. SIGNIFICANCE OF THE STUDY

The proposed framework offers several practical contributions:

- Enables early stress detection in organizational environments
- Supports data-driven mental health policies
- Preserves employee confidentiality
- Reduces dependency on manual screening
- Adaptable to different industries

VII. APPLICATIONS

This system can support several real life applications:

- Employee mental health monitoring systems
- Organizational well-being analytics
- Human resource decision support tools
- Workplace stress prevention programs
- Mental health research studies

Such systems can help organizations identify potential risks and implement appropriate interventions.

VIII. FUTURE DIRECTIONS

Upcoming research could enhance the suggested framework through various improvements. Sophisticated methods like deep learning, such as artificial neural networks, could help detect more intricate patterns within the dataset.

Integration with wearable devices could enable real-time monitoring of physiological indicators related to stress. Additionally, natural language processing techniques may be used to analyse employee feedback or communication patterns.

Developing web-based or mobile applications based on the proposed system could further improve accessibility and usability.

Future enhancements may include:

- Integration of deep neural networks for improved feature abstraction
- Incorporation of behavioral data from wearable devices
- Deployment as a cloud-based mental health monitoring system

IX. CONCLUSION

This study presented an ensemble machine learning framework for early prediction of occupational psychological stress. By leveraging structured survey data and supervised classification models, the system effectively identifies stress levels among employees.

Of the methods evaluated, Random Forest delivered the strongest predictive results, reaching an accuracy of 89.45%. These outcomes demonstrate that machine learning can effectively aid in the early detection and handling of mental health challenges in workplace settings.

The proposed approach contributes toward scalable, automated, and privacy-aware workplace well-being systems.

X.REFERENCES

1. Mental Disorder Detection: Bipolar Disorder Scrutinization using Machine Learning, published in 2019.
2. Intelligent data mining and machine learning for mental health diagnosis using genetic algorithm Azar, Ghassan &Gloster, Clay & El Bathy, Naser & Yu, Su &Neela, Rajasree & Alothman, Israa. (2015). Intelligent data mining and machine learning for mental health diagnosis using genetic algorithm. 201-206. 10.1109/EIT.2015.7293425
3. A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression B. Saha, T. Nguyen, D. Phung and S. Venkatesh, "A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression," in IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 4, pp. 1008 1015, July 2016.
4. Detecting Cognitive Distortions Through Machine Learning Text AnalyticsT. Simms, C. Ramstedt, M. Rich, M. Richards, T. Martinez and C. Giraud-Carrier, "Detecting

Cognitive Distortions Through Machine Learning Text Analytics," 2017 IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, 2017, pp. 508-512.

5. Machine Learning Framework for the Detection of Mental Stress at Multiple Levels Subhani, Ahmad & Mumtaz, Wajid & MOHAMAD SAAD, MOHAMAD NAUFAL & Kamel, Nidal& Malik, Aamir. (2017). Machine Learning Framework for the Detection of Mental Stress at Multiple Levels. 10.1109/ACCESS.2017.2723622. IEEE Access. PP. 1-1.