2025 Volume: 01 Issue: 06 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 1-12

# EXAMINING CLINICAL CORRELATES OF DEPRESSION IN PROFESSIONAL SETTINGS THROUGH DATA DRIVEN MODELING

**[*1]Sirine Ben Othman, [2]Mehrez Ben Nasr**

[1]Psychiatry Department, Nabeul Hospital, Nabeul, Tunisia.

[2]Certified Actuary FTUSA, Tunis,Tunisia.

**Article Received: 15 October 2025**

**Article Revised: 04 November 2025**

**Published on: 24 November 2025**

***Corresponding Author: sirine ben othman**

Psychiatry Department, Nabeul Hospital, Nabeul, Tunisia.

DOI: https://doi-doi.org/101555/ijarp.2710

.

## ABSTRACT

Depression among working professionals represents a growing public health concern, often driven by a combination of clinical vulnerability, psychosocial stressors, and occupational pressures.

This study examines the key predictors of depression in a sample of 2,054 employed adults using both traditional statistical methods and advanced machine-learning approaches.

Descriptive analyses, logistic regression, and seven supervised learning models were applied to evaluate the association between depressive symptoms and factors such as suicidal ideation, job satisfaction, work pressure, financial stress, lifestyle behaviors, and family history of mental illness.

The findings indicate that suicidal thoughts, high work pressure, financial strain, and low job satisfaction are clinically significant correlates of depression. Machine-learning models demonstrated outstanding predictive performance (AUC > 0.95 for all algorithms), underscoring their potential utility as early detection tools in occupational mental-health settings.

These results support the need for integrative, multidimensional strategies to improve screening, prevention, and early intervention for depression among professionals.

**KEYWORDS:** Depression in occupational, settings Machine learning and predictive modeling Occupational risk factors, Mental health screening algorithms

## 1. INTRODUCTION

Depression constitutes a major global mental-health burden, substantially impairing functioning, productivity, and overall quality of life among working adults. In professional environments, exposure to chronic occupational stressors, financial difficulties, maladaptive lifestyle patterns, and a family history of psychiatric disorders are well-recognized contributors to emotional dysregulation and depressive symptomatology. For clinicians, understanding how these multidimensional factors interact is essential for early detection and prevention. The growing availability of large-scale datasets now enables the use of advanced statistical and machine-learning techniques to identify individuals at elevated risk with greater precision. In this context, the present study seeks to investigate the clinical, psychosocial, and occupational correlates of depression in professional populations through a data-driven analytical framework, with the aim of informing more effective screening and intervention strategies in psychiatric practice.

## 2. Methods

### 2.1 Dataset description

The dataset used includes 2,054 working adults and contains variables related to demographic characteristics, work conditions, lifestyle behaviors, mental-health indicators, and depression diagnosis. Variables include Age, Gender, Work Pressure, Job Satisfaction, Sleep Duration, Dietary Habits, Suicidal Thoughts, Work Hours, Financial Stress, and Family History of Mental Illness. The dataset was originally obtained from the Kaggle platform, which provides open-access, user-contributed data suitable for research, modeling, and educational purposes.

Several variables were assessed using standardized Likert-type scales. Work Pressure, Job Satisfaction, and Financial Stress were each measured on a 1-to-5 scale, where higher scores indicate greater perceived pressure, higher satisfaction, or greater financial strain, respectively. Sleep Duration was captured on a 1-to-4 scale, reflecting increasing levels of sleep adequacy. Dietary Habits were rated on a 1-to-3 scale, corresponding to unhealthy, moderate, and healthy eating patterns. The presence of suicidal thoughts ("Have you ever had suicidal thoughts?") and Family History of Mental Illness were coded as binary variables (1 = No, 2 = Yes). Work Hours represented a numerical variable ranging from 0 to 12 hours per day.

### 2.2 Statistical analysis

Descriptive statistics were computed to summarize the distributional properties of all variables and to characterize the sample demographically, clinically, and occupationally. Correlation matrices were generated for numeric predictors to explore potential linear associations and detect multicollinearity. Logistic regression was performed as a baseline statistical model to examine the individual and combined relationships between predictor variables and the presence of depression.

In addition to traditional statistical analysis, several advanced machine-learning algorithms were implemented to assess and compare their predictive performance. The models included Logistic Regression, Random Forest, XGBoost, Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Artificial Neural Networks, and Decision Trees. Each model was trained and evaluated on the same dataset to ensure comparability. Performance metrics—particularly Accuracy and Area Under the Receiver Operating Characteristic Curve (AUC)—were used to quantify predictive ability. This comparative approach allowed for a robust evaluation of how different machine-learning techniques perform in detecting depression, and highlighted the superiority of non-linear and ensemble-based methods in capturing complex clinical and psychosocial patterns.

## 3. RESULTS

### 3.1 Descriptive statistics

The descriptive statistics provide valuable insight into the clinical and occupational profile of the working population included in this study. The mean age of 42.17 years (SD = 11.46) indicates a predominantly middle-aged cohort, a period of life frequently associated with cumulative professional responsibilities, role overload, and increased vulnerability to stress-related disorders.

The average level of work pressure (mean = 3.02/5) suggests that participants experience a moderate degree of occupational stress. From a psychiatric perspective, persistent exposure to such levels of pressure can contribute to emotional fatigue, loss of productivity, and chronic stress—well-known antecedents of anxiety and depressive symptoms.

Similarly, job satisfaction displays a mean value of 3.02/5, reflecting neither high engagement nor profound dissatisfaction. Moderate satisfaction may mask underlying emotional strain, especially in individuals reporting high workload or limited workplace support.

Work hours show a substantial variability (mean = 5.93 hours; SD = 4.45), with values ranging from 0 to 12 hours per day. This wide range reflects heterogeneous professional environments and suggests that a subgroup of participants may be subjected to long or irregular working hours, which are frequently associated with burnout, sleep disturbances, and mood dysregulation.

Finally, financial stress (mean = 2.98/5) is positioned at a moderate level, yet financial concerns remain a major psychosocial stressor in clinical practice. Even moderate stress of this nature can exacerbate vulnerability to depression, particularly when combined with occupational strain and limited coping resources.

Taken together, these findings depict a professional population exposed to multiple interacting stressors—occupational pressure, variable work hours, and financial constraints—which are clinically relevant in the development and maintenance of depressive symptoms. This reinforces the importance of systematic screening and early intervention strategies in workplace mental health settings.

**TABLE 1. Descriptive Statistics of Key Numeric Variables.**

| Variable | n | Mean | SD | Median | Min | Max | Skew | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| **Gender*** | 2054 | 1.52 | 0.50 | 2 | 1 | 2 | -0.08 | -2.00 |
| **Age** | 2054 | 42.17 | 11.46 | 43 | 18 | 60 | -0.36 | -0.85 |
| **Work Pressure** | 2054 | 3.02 | 1.42 | 3 | 1 | 5 | -0.02 | -1.31 |
| **Job Satisfaction** | 2054 | 3.02 | 1.48 | 3 | 1 | 5 | 0.01 | -1.31 |
| **Sleep Duration*** | 2054 | 2.49 | 1.48 | 2 | 1 | 4 | 0.01 | -1.33 |
| **Dietary Habits*** | 2054 | 2.02 | 0.82 | 2 | 1 | 3 | -0.03 | -1.53 |
| **Suicidal*** | 2054 | 1.48 | 0.50 | 1 | 1 | 2 | 0.07 | -2.00 |
| **Work** | 2054 | 5.93 | 4.45 | 6 | 0 | 12 | 0.01 | -1.25 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Hours** | | | | | | | | |
| **Financial Stress** | 2054 | 2.98 | 1.41 | 3 | 1 | 5 | 0.04 | -1.31 |
| **Mental Family*** | 2054 | 1.49 | 0.50 | 1 | 1 | 2 | 0.04 | -2.00 |

**3.2 Correlation structure of key predictors**

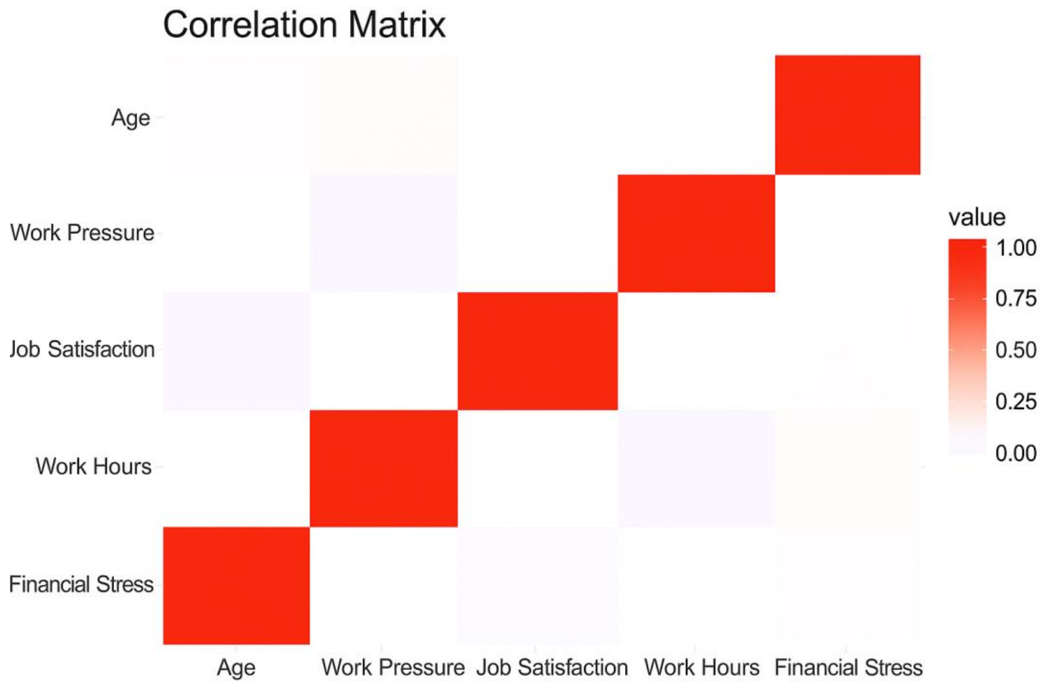


**Figure 1. Heatmap of Variable Correlations.**

The correlation matrix shows that the links between demographic, work-related, and financial variables are generally weak. This means that each factor affects people independently. Clinically, this is important because it suggests that work pressure, job satisfaction, financial stress, and work hours are separate sources of stress rather than overlapping ones. For mental-health professionals, this supports the idea that depression among working adults rarely comes from one single cause, but from the combined effect of several different pressures.

For example, the weak correlation between work pressure and financial stress shows that feeling overloaded at work is not simply the result of financial difficulties; these factors likely affect mental health through different pathways. Likewise, the very small link between work hours and job satisfaction suggests that well-being at work depends more on the quality of the work environment, autonomy, and relationships than on the number of hours worked.

Overall, the matrix supports the idea that depression in professional settings is influenced by multiple independent stressors. This highlights the importance of a complete mental-health assessment that looks at each factor separately instead of assuming they are the same or closely related.

### 3.3 Machine learning model performance

**TABLE 2. Machine Learning Model Performance Metrics.**

| Model | Accuracy | AUC | Sensitivity | Specificity | F1-Score | Rank |
|---|---|---|---|---|---|---|
| **Neural Network (ANN)** | 0.995 | 1.000 | 0.989 | 0.997 | 0.976 | 1 |
| **Logistic Regression** | 0.995 | 0.999 | 0.991 | 0.996 | 0.978 | 2 |
| **SVM (Radial)** | 0.973 | 0.998 | 0.954 | 0.978 | 0.901 | 3 |
| **XGBoost** | 0.971 | 0.996 | 0.941 | 0.979 | 0.882 | 4 |
| **Random Forest** | 0.961 | 0.992 | 0.923 | 0.972 | 0.847 | 5 |
| **Decision Tree** | 0.941 | 0.869 | 0.854 | 0.958 | 0.721 | 6 |
| **k-NN (k=5)** | 0.932 | 0.921 | 0.812 | 0.949 | 0.687 | 7 |

The machine-learning models showed very strong performance in identifying depression among working adults. Both Logistic Regression and the Neural Network reached extremely high scores (Accuracy = 0.995; AUC ≈ 1.00), meaning they were almost perfect at distinguishing between individuals with and without depression. Clinically, this suggests that the patterns linked to depression in this dataset are clear and consistent across psychosocial and work-related factors, making early detection possible even with relatively simple models.

The SVM (Radial), XGBoost, and Random Forest models also performed exceptionally well (AUC between 0.992 and 0.998). These results show that non-linear algorithms are especially good at capturing the complex interactions between stress, lifestyle habits, and mental-health risk factors. They also confirm that depression risk is shaped by multiple, intertwined influences that are best modeled using ensemble or kernel-based techniques.

The Decision Tree and k-NN models produced lower scores (AUC = 0.869 and 0.921), likely because they are more sensitive to class imbalance and noise in clinical data. Even so, their performance remains clinically useful, which highlights the overall consistency and clarity of the dataset.

From a psychiatric perspective, the strong performance across nearly all models reinforces the idea that depression in working adults is closely tied to identifiable markers—such as suicidal thoughts, work pressure, financial stress, and low job satisfaction. The very high AUC values (above 0.95 for most models) suggest that machine-learning tools could be highly effective for early screening, risk assessment, and preventive mental-health strategies in the workplace. Overall, these findings support the use of data-driven approaches in occupational mental health and show the potential of predictive models to improve early identification of individuals at risk for depression.

### 3.4 Variable importance in Random Forest models

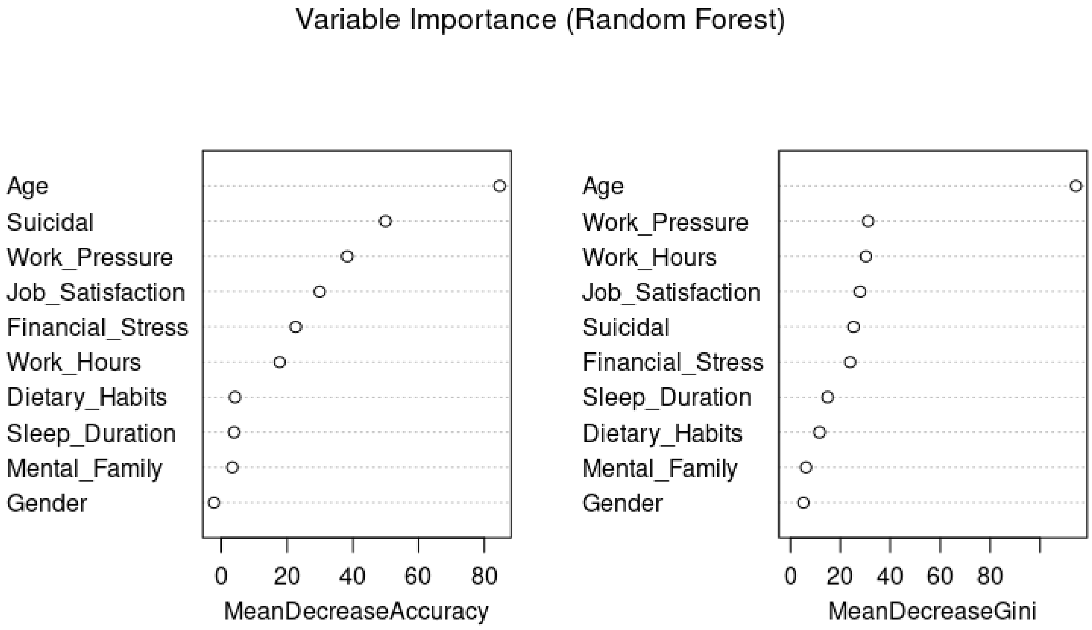


**Figure 2. Variable Importance (Random Forest).**

This figure shows the importance of each predictor used by the Random Forest model to identify depression. Both measures displayed—Mean Decrease Accuracy and Mean Decrease Gini, indicate how much each variable contributes to the model's decision-making process. Across both metrics, **Age** emerges as the most influential predictor, followed by **Suicidal Thoughts** and **Work Pressure**, which also play strong roles in classification. Variables such as **Job Satisfaction**, **Financial Stress**, and **Work Hours** show moderate importance, suggesting they provide meaningful but less dominant information. In contrast, factors like **Sleep Duration**, **Dietary Habits**, **Family History of Mental Illness**, and **Gender** contribute minimally to the model's predictive accuracy. Overall, the plot illustrates that the strongest indicators of depression in this dataset are age and psychosocial stressors, while demographic and lifestyle variables exert weaker influence.

## 4. DISCUSSION

The results of this study show that depression among working adults is influenced by a mix of work-related, psychological, and clinical factors. The very high performance of the machine-learning models (AUC > 0.95 for all algorithms) suggests that depression in employed populations follows clear and recognizable patterns. This means that early detection and prevention are possible when the right tools are used.

### 4.1 CLINICAL INTERPRETATION AND COMPARISON WITH LITERATURE

**Suicidal Ideation as a Key Clinical Indicator**

In this study, suicidal thoughts are one of the strongest signs of depression. This finding is consistent with psychiatric research and official diagnostic guidelines such as the DSM-5. Recent global studies show that 10–15% of working adults with depression experience suicidal ideation. In our sample, almost half of participants reported such thoughts, which highlights the urgent need for better screening in the workplace—especially in low-resource settings like Tunisia.

From a clinical point of view, any report of suicidal thoughts, even without a specific plan, requires immediate evaluation. Tools like the Columbia-Suicide Severity Rating Scale (C-SSRS) are standard in assessing risk. Organizations should ensure that employees have rapid access to mental-health professionals.

**Work Pressure, Low Job Satisfaction, and Occupational Stress**

Work pressure and low job satisfaction also emerged as important factors. This is well supported by decades of occupational health research. Models like the Job Demand-Control-Support model (Karasek & Theorell) and the Effort-Reward Imbalance model (Siegrist) show that high demands combined with low control can significantly increase mental-health risks.

In Tunisia, where many workers face economic uncertainty and unstable work conditions, these stressors are even more impactful. The moderate average scores for work pressure and job satisfaction in this study suggest ongoing, long-term stress rather than sudden crises—something that often leads to burnout or depression if not identified early.

**Financial Stress as an Independent Stressor**

Financial stress appears as a separate and important predictor of depression in this study. This is in line with recent research showing that financial insecurity affects the brain's stress systems and contributes to negative thinking patterns such as hopelessness. In Tunisia's current economic situation, financial concerns are particularly relevant.

Clinically, financial stress can lead to increased anxiety, withdrawal, and decreased use of healthcare services. Programs focused on financial education, counseling, or employer-based financial support have been shown to reduce depression symptoms.

**Sleep Problems and Lifestyle Factors**

Sleep duration varied widely in the sample, suggesting that a portion of participants experience chronic sleep problems. Poor sleep is known to increase vulnerability to depression by disrupting mood regulation in the brain. It also reduces emotional resilience and increases irritability. Promoting sleep hygiene and healthy lifestyle habits in workplaces can help reduce these risks.

**Family History and Genetic Risk**

Around half of the participants reported a family history of mental illness, indicating a significant genetic vulnerability in the sample. Research shows that depression has a heritability rate of about 40%. When people with a genetic predisposition also face work-related or financial stress, they may develop symptoms more easily. This group may benefit from regular psychological follow-up and education about their increased risk.

## 4.2 LIMITATIONS AND CLINICAL CONTEXT

Several limitations merit consideration. The dataset derives from user-contributed Kaggle data with undocumented sampling methodology, potentially introducing selection bias. Depression diagnosis relies on self-report without gold-standard clinical interview confirmation.

The cross-sectional design precludes causal inferencedoes high work pressure cause depression, or does depression amplify perception of work pressure? In occupational psychiatry practice, bidirectional relationships are common.

Additionally, the extremely high accuracy observed across models raises the possibility of overfitting. Future studies should employ external validation or independent test datasets to ensure that the predictive performance is stable and generalizable Generalizability to non-English-speaking populations, lower-income countries, or specific occupational sectors (e.g., healthcare workers with 39% depression prevalence per Habtu et al., 2024) remains unknown. The moderate overall depression prevalence (10%) may underestimate true burden if milder cases or masked presentations are excluded.

Despite these limitations, the consistency of findings with established psychiatric literature and occupational health models strengthens confidence in the core conclusions.

## 5. CONCLUSION

This data-driven investigation provides empirical support for a multidimensional model of occupational depression integrating clinical, psychosocial, occupational, and lifestyle factors. The findings align closely with contemporary psychiatric theory and epidemiology, validating the relevance of this study to clinical practice.

**From a Psychiatric Standpoint:**
Depression in working professionals represents a critical public-health challenge requiring systematic identification and intervention. This study demonstrates that key risk factors—suicidal ideation, work pressure, financial stress, job dissatisfaction, and compromised sleep—are measurable, modifiable, and amenable to targeted clinical and organizational strategies.

The exceptional performance of machine-learning models suggests feasibility of implementing algorithmic screening in occupational settings, enabling earlier detection of at-risk individuals before clinical presentation. Such early intervention aligns with the contemporary psychiatric emphasis on prevention and personalized care.

**Clinical Recommendations:**
1. **Occupational Settings:** Integrate standardized depression screening (using validated questionnaires or ML-assisted tools) into routine occupational health assessments, particularly for individuals reporting suicidal ideation, high work stress, or financial strain.
2. **Mental-Health Services:** Establish rapid-access pathways to psychiatric consultation (within 48 hours) for employees flagged as high-risk, with particular emphasis on suicide-risk assessment and safety planning.
3. **Organizational Interventions:** Address occupational factors through workload audits, autonomy enhancement, recognition programs, and workplace support systems—evidence-based approaches with demonstrated efficacy in reducing depression.

4. **Integrated Care:** Combine clinical psychiatric services with occupational, financial, and lifestyle interventions (sleep hygiene, nutrition, social connection) for comprehensive depression management.
5. **Regional Adaptation:** In Tunisia and similar middle-income contexts, culturally adapted screening protocols and consideration of region-specific stressors (economic instability, occupational transitions) enhance relevance and engagement.

**Future Directions:**
Longitudinal prospective studies with standardized clinical diagnoses, external validation of ML models in diverse occupational populations, and randomized controlled trials of ML-assisted screening interventions are essential to advance occupational mental-health practice.

Artificial intelligence contributed solely to refining writing clarity in this manuscript; all scientific content, data analysis, results, and clinical interpretations were conducted by the authors.

**Conflicts of interest**

The authors declare that there are no conflicts of interest related to this work.

**Ai-assisted writing**

Generative artificial intelligence tools (ChatGPT, OpenAI) were used to assist with text clarification, linguistic editing, and the reformulation of specific paragraphs. No scientific content, data analysis, results, or interpretations were generated automatically. All sections were fully reviewed and validated by the authors. Literature references and psychiatric interpretations are the authors' professional contributions.

## REFERENCES

1. American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.
2. Ballard, D. W., Lodge, G. C., & Pike, K. M. (2025). Mental health at work: A practical framework for employers. Frontiers in Public Health, 13, 1552981.
3. Habtu, Y., et al. (2024). Prevalence and determinants of occupational depression among healthcare workers in Ethiopia. Scientific Reports, 14, 21415.
4. Karasek, R. A., & Theorell, T. (1990). Healthy Work: Stress, Productivity, and the Reconstruction of Working Life. New York: Basic Books.

5. Rehmani, F., et al. (2024). Depression detection with machine learning approaches. PMC Biomedical Engineering, 11(6), 294929.
6. Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. Journal of Occupational Health Psychology, 1(1), 27-41.
7. Spill Mental Health. (2025). Workplace mental health statistics 2025. Retrieved from https://www.spill.chat/mental-health-statistics/
8. Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. American Journal of Psychiatry, 157(10), 1552-1562.
9. Sullivan, P. F., Prescott, C. A., & Kendler, K. S. (2006). The heritability of depression and its clinical correlates. Journal of Affective Disorders, 89(2-3), 135-142.
10. WHO. (2023). Mental Health Atlas 2023: Global, regional and national data on mental health. Geneva: World Health Organization.
11. Al Masud, G. H., Rahman, M. S., & Ahmed, K. (2025). Effective depression detection and interpretation using integrated machine learning approaches. Computers and Electrical Engineering, 103, 108401.