DOI: 10.54254/2755-2721/157/2025.PO23118

Identifying Risk Factors and Developing Predictive Models for Depression Using Machine Learning Analysis of Indian Survey Data

Yiyang Liu

The Webb Schools, Claremont, USA
stephanie_yyliu@163.com

Abstract: Understanding the correlation between various risk factors and depression is important for developing personalized and effective treatment plans, especially for those at risk of recurrent or chronic depression. Current literature and predictive models identify pressure and age as primary risk factors, in addition to other factors such as financial stress, employment or student status, and suicidal thoughts. This approach emphasizes the importance of considering a range of interrelated factors that may influence each other and compound over time. Particular attention should be paid to vulnerable and marginalized populations, who are disproportionately affected and have higher prevalence of depression, emphasizing the need for individualized treatment and targeted interventions. High accuracy predictive models are used to assess the impact of individual and combined risk factors, providing a tool for identifying people at risk of depression before symptoms such as somatization become fully manifest. Early identification can lead to more proactive interventions, improved clinical outcomes, and a reduction in the overall impact of depression on society.

Keywords: Depression, risk factors, mental health

1. Introduction

Mental health is an important aspect of a person’s overall well-being, affecting how they think, feel, and act. Good mental health contributes to a positive quality of life, enabling people to cope with challenges and manage stress. However, when mental health is compromised, the negative effects are felt not only by the individual, but also by their family, community, and society as a whole. Mental health issues are a growing concern in today’s society. According to SAMHSA's National Survey, 23% of adults experienced a mental illness in the year 2024, which is nearly 60 million Americans. [1]

Depression is one of the most recognized and common mental health disorders. Emotional symptoms include experiencing feelings of sadness, emptiness, hopelessness, and loss of interest in daily activities. Physical symptoms include, but are not limited to, changes in sleep patterns, appetite, tremors, pain, tinnitus, and difficulty concentrating or remembering. However, the diagnosis of depression is complex. There is ongoing debate about its precise definition, boundaries, and classification, making consistent diagnosis difficult. Cultural factors further complicate the issue, as different cultures interpret and diagnose depression based on their own perceptions of the

DOI: 10.54254/2755-2721/157/2025.PO23118

symptoms.[2] In addition, situational factors, such as seasonal changes or social pressures, can cause normal distress to be misclassified as depression.[3] This uncertainty applies to intervention strategies for which there is a lack of conclusive evidence to determine their effectiveness for both clinical and non-clinical depressed populations. The mechanisms responsible for reducing depressive symptoms remain unclear as well.

The impact of depression extends beyond the individual. For individuals, depression often means health care costs, especially if the condition leads to long-term disability. There is also a social stigma that tends to dismiss depression, which can prevent individuals from seeking the help they need. Under these circumstances, 77% of adults experiencing mental health issues did not receive treatment in 2024, especially those from racial and ethnic minorities.[1] Families and communities are also be burdened by the emotional and financial stress that depression can cause. On a larger scale, high rates of depression can lead to social isolation, reduced community engagement, and lost productivity, all of which have a negative impact on the economy.

The identification of depression risk factors serves dual purposes: understanding the underlying causes of depression and guiding targeted intervention strategies. For example, depression screening has shown that positive results are more common among individuals with lower incomes, public health insurance, less than a high school education, or who are separated, divorced, or widowed.[4] These sociodemographic factors, along with biological and psychological variables, create unique risk profiles that can significantly influence both vulnerability to depression and response to treatment. Early intervention, when implemented systematically, is effective in preventing or delaying mental disorders and offers accessible, cost-effective, and positive outcomes. For example, exercise has been shown to be an effective intervention for people with mild to moderate depression, showing moderate to large antidepressant effects and even greater efficacy for those with clinically diagnosed depression.[5] In addition, ongoing medication and preventive psychotherapy show strong benefits for relapse prevention.

Thus, understanding the correlation between possible risk factors and depression is crucial for developing personalized and effective treatment plans, especially for those at risk of recurrence or chronic depression. Despite extensive research on depression treatments, less attention has been given to how specific risk factors might influence the effectiveness of different intervention approaches. By identifying the most significant risk factors, this research aims to develop a predictive model that could help clinicians identify individuals at heightened risk for depression before clinical symptoms fully manifest, allowing for earlier and potentially more effective interventions.

2. Literature review

A comprehensive literature review was conducted to explore current research on classification and characteristics of common mental health issues in modern society, with a specific focus on depression risk factors and interventions.

Mofatteh identified mental health risk factors in several areas: Psychological factors include personality traits, underlying mental illness, and coping mechanism; environmental factors include high pressure, long work hours, and inadequate financial support; lifestyle factors include stimulant consumption, poor habits, and insufficient sleep; social factors include relationships, minority status, and stigma associated with mental health.[6] Hammen also identified biological factors such as having depressed parents and gender [7]. These categories provide a useful framework for understanding the multifaceted nature of depression risk, with particular emphasis on how various factors interact and compound over time. While specific prevalence rates differ across populations, these risk factors collectively create vulnerability profiles that can help identify individuals at heightened risk before clinical symptoms manifest. While the prevalence of specific risk factors

DOI: 10.54254/2755-2721/157/2025.PO23118

remains unclear, they have been shown to play a significant role in the diagnosing depression and determining the effectiveness of interventions.

Several studies have examined the prevalence of depression in relation to different sociodemographic factors and identified their associations with risk factors. Olfson et al. found that depression was more common among individuals with lower income, public health insurance, less than a high school education, or who were separated, divorced, or widowed.[4] Similarly, Ramón-Arbués et al. found a higher prevalence of depression among female college students.[8] Wang et al. reported a 27.0% prevalence of depression or depressive symptoms among outpatients, with the highest rates found in otolaryngology clinics. They also found that depression was more prevalent in developing countries than in developed countries, especially among those aged 80 years or older, likely due to factors such as physical dysfunction, low personal control, and loss of status.[9] All studies show that depression is more prevalent among marginalized or disadvantaged groups, who often face additional challenges such as limited social support or lower social status. This finding emphasizes the need for individualized treatment and targeted attention to these vulnerable populations.

Several studies have examined the effectiveness of various early interventions for depression. McGorry and Mei found that early intervention is important to preventing or delaying mental disorders and offers positive, accessible, and cost-effective outcomes when implemented systematically.[10] Specifically, Josefsson et al. identified exercise as an effective intervention for mild to moderate depression, with benefits such as stress reduction, improved cardiovascular health, and improved cognitive function.[5] Their research particularly highlighted how exercise interventions can be tailored to individual risk profiles, making them an important component to consider in a predictive framework for depression management. While Ormel et al. found that medication and preventive psychotherapy are highly effective for preventing relapse.[11] Their work emphasizes the importance of considering long-term outcomes and relapse prevention when developing predictive models for depression treatment, suggesting that initial risk factors may continue to influence treatment response over extended periods.

Research on depression interventions in specific settings and populations has revealed important contextual factors that influence both risk profiles and treatment efficacy. Fazel et al. identified the benefits of integrating mental health services into education systems to improve physical health, mental health and educational outcomes, although the effectiveness is limited by insufficient prioritization of child and adolescent health.[12]

This setting-specific approach underscores how environmental contexts modify both risk presentation and intervention effectiveness, suggesting that predictive models must account for these contextual variables to accurately forecast depression risk and treatment response.

The existing literature presents several notable limitations that our current research aims to address. Such limitations include gaps in understanding the mechanisms behind reductions in depressive symptoms, uncertainty about the long-term effectiveness of different treatment methods, difficulties with generalization, and underrepresentation of certain groups in data collection. Particularly relevant to our predictive modeling approach is the lack of integration between risk factor identification and treatment response prediction. While studies like Mofatteh's have identified risk categories, and others like Josefsson et al. have evaluated specific interventions, there remains insufficient research connecting these two areas into cohesive predictive frameworks. These findings emphasize the importance of further research to identify risk factors and refine intervention strategies to improve mental health outcomes.

DOI: 10.54254/2755-2721/157/2025.PO23118

3. Sources of data

Data were analyzed from an anonymous survey conducted between January and June 2023 in various cities in India, targeting adults between the ages of 18 and 60. The survey aimed to identify the risk factors associated with depression. Participants provided information on a range of demographic and lifestyle factors without the need for a professional mental health assessment or diagnostic test results.

The target variable, depression, was defined based on participants' responses indicating whether they were at risk for depression. The dataset is hosted on Kaggle, containing 2,556 entries and 19 columns. It includes a variety of risk factors, including demographic, academic or work, physical, and psychological aspects. This diversity allows for a comprehensive analysis of the variables that may contribute to an individual's risk for depression, providing insights for targeted intervention strategies.

4. Exploratory data analysis

Exploratory data analysis (EDA) was performed on the dataset to summarize its key characteristics, uncover patterns, and identify irregularities. Variables were categorized as numeric (discrete or continuous) and categorical (nominal or ordinal). The data were summarized using descriptive statistics such as mean, median, mode, and standard deviation to provide insight into central tendencies and variability. Multiple visualization techniques were employed, including bar graphs, pie charts, and box plots, to help identify distributions and relationships between variables.

In addition, the analysis distinguished between prognostic factors that correlate with depression (such as gender, age, and city) and predictive factors that may have causal relationships with depression (such as work pressure, sleep habits, and financial stress). This comprehensive approach to EDA guided the subsequent analysis and modeling phases by identifying the most promising variables for predictive modeling.

The exploratory analysis revealed several significant relationships between various factors and depression:

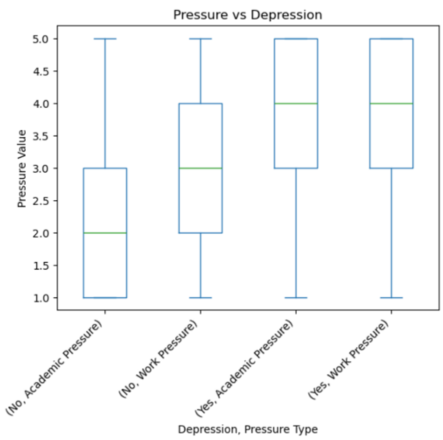
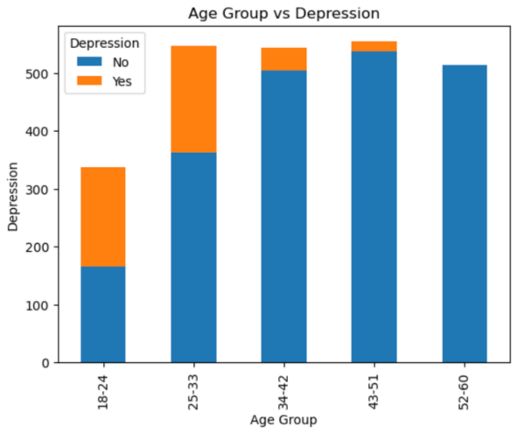


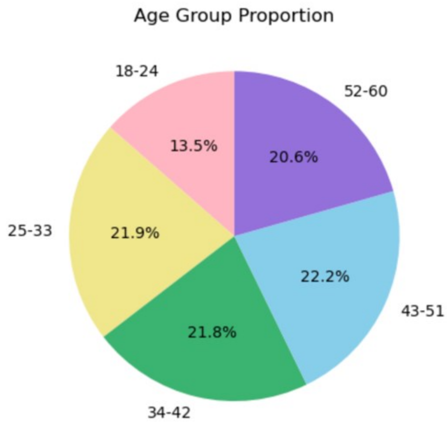
Figure 1: Comparison of pressure levels in depressed and non-depressed individuals

Pressure and Depression: The box plot comparison shows that individuals with depression experience higher levels of pressure (median and third quartile values), suggesting a positive correlation between pressure intensity and depression risk.

DOI: 10.54254/2755-2721/157/2025.PO23118



(a) Depression prevalence by age group



(b) Age group distribution

Figure 2: Analysis of the prevalence and distribution between age and depression

Age and Depression: The box plot shows that depression is most common among young adults (18-24), with nearly half experiencing it, followed by about one-third of the 25-33 age group. The prevalence decreases with age, affecting only 5% of the 34-42 age group, 3% of the 43-51 age group, and becoming almost negligible in the 52-60 age group. This trend suggests a negative correlation between age and depression. The pie chart shows a relatively balanced distribution between each age group, with the exception of the 18-24 age group, with represents 13.5% of the total. This imbalance may affect the interpretation of the proportion of depressed individuals relative to the size of the group.

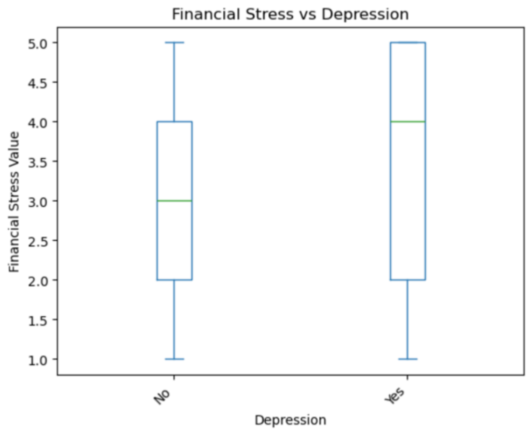
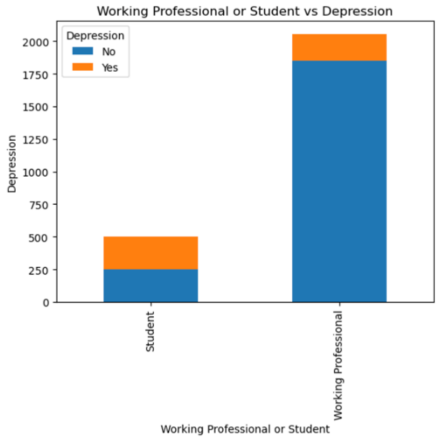


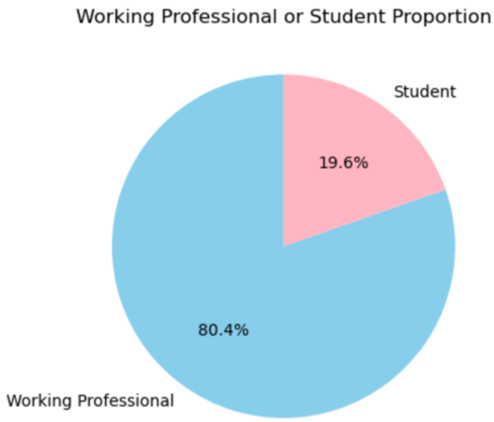
Figure 3: Comparison of financial stress levels in depressed and non-depressed individuals

Financial Stress and Depression: Box plot analysis reveals higher median and third quartile values of financial stress among individuals with depression, indicating a positive correlation between financial difficulties and depression risk.

DOI: 10.54254/2755-2721/157/2025.PO23118



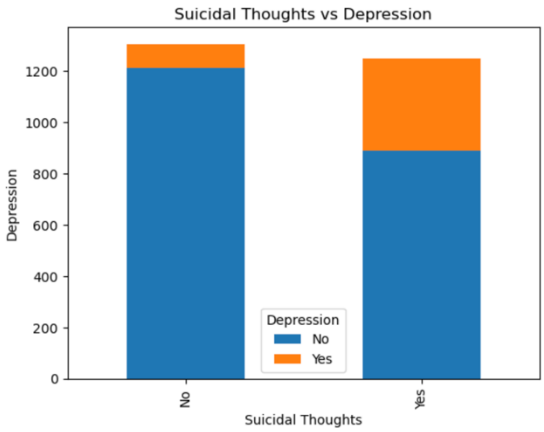
(a) Depression prevalence by employment or student status



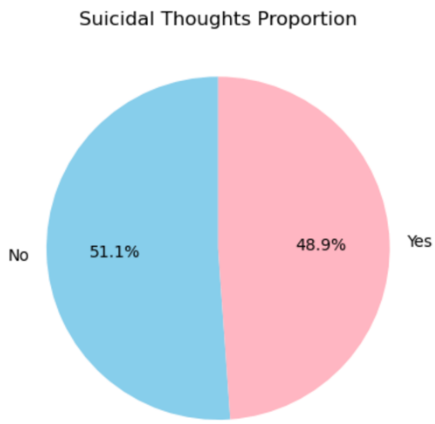
(b) Employment or student status distribution

Figure 4: Analysis of the prevalence and distribution between employment or student status and depression

Academic and Work Pressure: Comparative box plots demonstrate that individuals without depression typically report academic pressure with a median of 2 (mostly between 1-3) and work pressure with a median of 3 (mostly between 2-4). In contrast, those with depression report both academic and work pressure with a median of 4 (mostly between 3-5), demonstrating elevated pressure levels across both domains for depressed individuals.



(a) Depression prevalence by suicidal thoughts



(b) Suicidal thoughts distribution

Figure 5: Analysis of the prevalence and distribution between suicidal thoughts and depression

Suicidal Thoughts: Depression is substantially more prevalent (approximately 30%) among those reporting suicidal thoughts compared to those without such thoughts (approximately 8%). The population studied shows a relatively balanced distribution between those with and without suicidal thoughts.

DOI: 10.54254/2755-2721/157/2025.PO23118

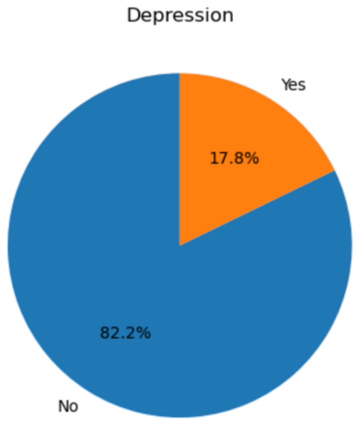


Figure 6: Depression distribution

Depression Prevalence: The overall sample shows an imbalanced distribution, with 82.2% classified as non-depressed and 17.8% as depressed. This imbalance is important to consider when interpreting findings and developing predictive models.

5. Methods

The dataset is imbalanced, with 17.8% depressed and 82.2% non-depressed. This trend reflects real-world scenarios where the prevalence of depression is relatively low compared to the non-depressed population. However, this imbalance in the dataset can affect the performance of machine learning models, as they will perform better with the majority class. Accuracy is also less informative, as a model that makes all non-depressed predictions will still achieve 82.2% accuracy. Three different methods are applied to address class imbalance in the dataset. The original dataset is split into a training set and a testing set, with 80% of the data used for training and 20% of the data used for testing. The training set is then balanced using each method. A logistic regression model is trained on each balanced training set to predict depression. Finally, the X values from the testing set are inputted into the trained model to generate predictions, which are compared to the actual outputs in the testing set. This comparison evaluates the performance of the model.

5.1. Data preprocessing and handling class imbalance

Three different methods — two types of oversampling and SMOTE — are applied to address class imbalance in the dataset. Oversampling duplicates random samples in the minority class with replacement to make the number of samples in the minority class equal to the number of samples in the majority class. The difference between the two types of oversampling is that one is manual and the other one automates the process. On the other hand, SMOTE generates synthetic samples for the minority class to make the number of samples in the minority class equal to the number of samples in the majority class. For each sample in the minority class, it randomly selects one of its five nearest neighbors, generating synthetic samples by picking a random point along the line connecting the sample to the selected nearest neighbor.

DOI: 10.54254/2755-2721/157/2025.PO23118

5.2. Model development and evaluation

5.2.1. Logistic regression

Logistic regression is used to make predictions about the dependent variable, depression, based on the independent variables, risk factors. The model uses the sigmoid function to predict the dependent variable, which ranges from 0 (not depressed) to 1 (depressed). The coefficients of the logistic regression represent the relationship between each feature and the target (depression). By analyzing these coefficients, we can assess the impact of multiple risk factors on the state of depression. A higher coefficient value indicates a stronger correlation between the risk factor and depression.

5.2.2. Random forest

The random forest algorithm is an ensemble machine learning algorithm that uses the Gini index to determine the importance of each feature. It is used to estimate the relative importance of each risk factor contributing to depression. First, a decision tree is trained on multiple randomly created subsets of the original dataset. A random subset of features is selected for each split in the decision trees, which helps to decorrelate the trees and increase the robustness of the ensemble. Each decision tree makes a prediction independently, so that the collective results of all trees reduce the risk of overfitting, thus improving the overall accuracy. Finally, majority voting is used to determine the final classification result.

5.3. Performance metrics

5.3.1. Confusion matrix

The performance of the model's predictions for each method used to address the imbalanced dataset is evaluated using a confusion matrix. The confusion organizes the predictions into four categories: a true positive, which is a correct prediction of depression; a false positive, which is an incorrect prediction of depression; a true negative, which is a correct prediction of non-depression; and a false negative, which is an incorrect prediction of non-depression.

5.3.2. Classification report

The performance of the model's predictions for each method used to address the imbalanced dataset is also evaluated using a classification report, which provides the calculated values of key measurements. Accuracy evaluates the overall performance of the model by calculating the percentage of correct predictions (true positives and true negatives) out of all predictions (true positives, false positives, true negatives, and false negatives). Precision focuses on the model's performance in predicting positives by calculating the percentage of true positives out of all predicted positives. Recall focuses on the model's the performance in identifying positives in the original dataset by calculating the percentage of true positives out of all actual positives (true positives and false negatives). The F1-score combines both precision and recall, providing a better assessment of the model's performance, especially in an imbalanced dataset such as this one. Given that the dataset consists of 17.8% depressed and 82.2% non-depressed, a model predicting an 100% non-depressed rate would have an accuracy of 82.2%. However, this model would be biased because a prediction of 100% non-depressed provides little meaningful insight or analysis about depression.

DOI: 10.54254/2755-2721/157/2025.PO23118

6. Results

6.1. Random forest feature importance

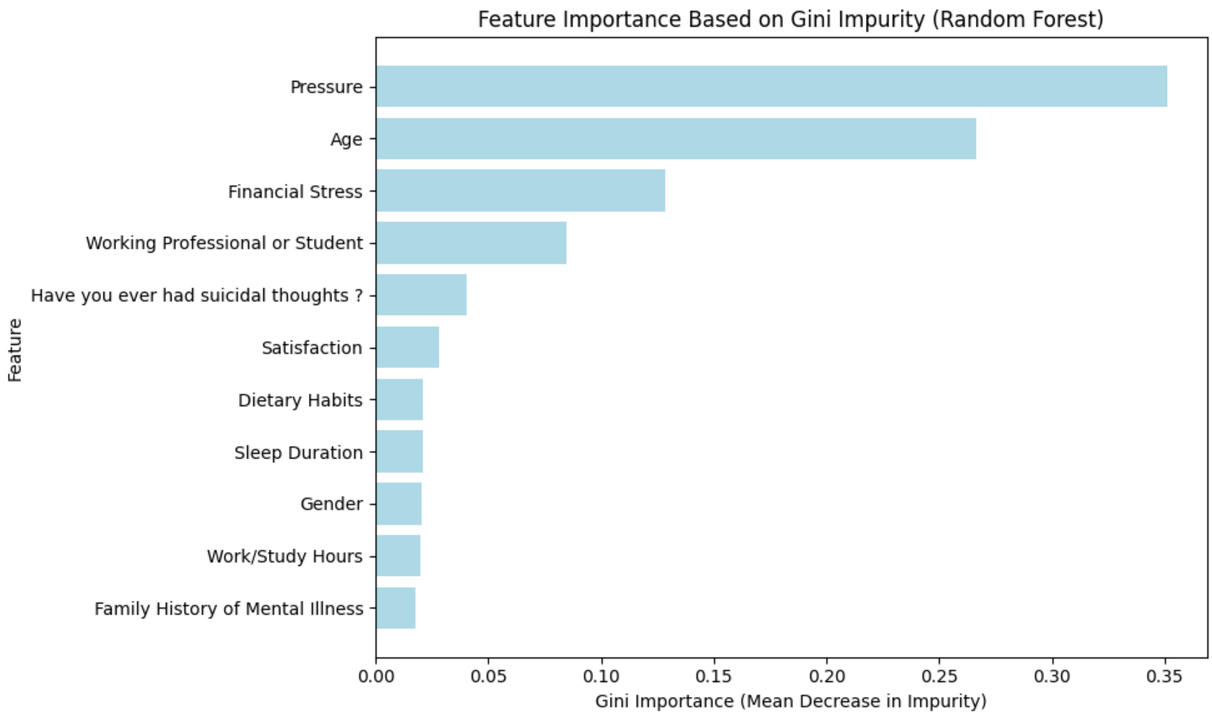


Figure 7: Importance of depression risk factors based on Gini Impurity

The random forest algorithm identified key predictors of depression through Gini importance measures, which indicate how much each feature contributes to reducing prediction uncertainty. Higher values signify greater predictive importance.

The graph shows the Gini importance values:

- Pressure: 0.35
- Age: 0.27
- Financial Stress: 0.13
- Working Professional or Student: 0.09
- Have you ever had suicidal thoughts?: 0.04

These results demonstrate that "pressure" and "age" are the most influential predictors of depression, followed by "financial stress," employment/student status, and suicidal thoughts. The remaining six features in the dataset showed minimal importance (Gini values of 0.01-0.03) and were excluded from subsequent modeling to improve efficiency.

To optimize model performance, we employed 5-fold cross-validation and hyperparameter tuning, which yielded the following optimal parameters:

- Number of trees: 200
- Maximum tree depth: None (no limit on the tree depth)
- Minimum samples required per split: 5
- Minimum samples required per leaf: 1
- Number of features considered per split: Square root of the total number of features

DOI: 10.54254/2755-2721/157/2025.PO23118

These hyperparameters help create a balanced model that improves generalization by adjusting complexity to capture underlying data patterns while avoiding overfitting.

6.2. Model performance comparison

Table 1: Comparison of model performance metrics for different sampling techniques

	Original	OverSampling 1	OverSampling 2	SMOTE
Confusion Matrix	390, 39 8, 75	389, 40 9, 74	389, 40 9, 74	390, 39 8, 75
Accuracy	0.91	0.90	0.90	0.91
Precision (1)	0.66	0.65	0.65	0.66
Recall (1)	0.90	0.89	0.89	0.90
F1-Score (1)	0.76	0.75	0.75	0.76

The confusion matrix and classification report were generated based on the inputs (risk factors) of “pressure”, “age”, “financial stress”, “working professional or student”, and “have you ever had suicidal thoughts?”. These five features were identified by the random forest model as the most influential factors for depression.

The original and SMOTE models have slightly more true positives and true negatives than the OverSampling 1 and OverSampling 2 models. As a result, the original and SMOTE models have slightly higher classification report values, with accuracy of 91%, precision of 66%, recall of 90%, and F1-Score of 76%. However, the differences in the classification report values between all models are minimal, indicating that their performance is fairly balanced. Since the dataset is imbalanced, the SMOTE technique likely provides more benefits than the original model by helping to balance the dataset and improve the performance of the minority class.

Table 2: Comparison of logistic regression model performance and coefficients for different sampling techniques

	Original	OverSampling 1	OverSampling 2	SMOTE
Pseudo R-squared	0.6453	0.6921	0.6916	0.6845
Log-Likelihood	-343.94	-713.67	-714.72	-731.34
LLR p-value	2.048e-268	0.000	0.000	0.000
Iterations	9	9	9	8
Constant	-3.5537	-2.2318	-2.2735	-1.3042
Pressure Coefficient	1.2401	1.3463	1.3497	1.4100
Age Coefficient	-0.2427	-0.2631	-0.2611	-0.2853
Financial Stress Coefficient	0.8310	0.9103	0.8962	0.8508
Employment/Student Status Coefficient	1.6815	1.5978	1.6166	1.1943
Suicidal Thoughts Coefficient	3.8617	4.2236	4.2356	4.1377

The logistic regression models were generated based on the inputs (risk factors) of “pressure”, “age”, “financial stress”, “working professional or student”, and “have you ever had suicidal thoughts?”. These five features were identified by the random forest model as the most influential factors for depression.

DOI: 10.54254/2755-2721/157/2025.PO23118

The pseudo R-squared value indicates how well the logistic regression model fits the data, with higher values indicating a better fit. In this case, the oversampling and SMOTE models have higher pseudo R-squared values compared to the original model, which means that balanced classes fit the model better. The log-likelihood value also assesses how well the logistic regression model fits the data, with higher values (less negative) indicating a better fit. The oversampling and SMOTE models have more negative log-likelihood values compared to the original model, likely due to the increased complexity of the models with class balancing techniques. The Likelihood Ratio Test (LLR) p-value tests whether the model improves the fit compared to the null model. A p-value greater than 0.05 indicates that the model does not significantly improve the fit and has little impact in terms of statistical significance. The original model has a p-value of 2.048e-268, indicating extremely strong evidence against the null hypothesis. The oversampling and SMOTE models have a p-value of 0.000, indicating even stronger evidence against the null hypothesis. Since 0.000 is smaller than 2.048e-268, the oversampling and SMOTE models improve the fit more compared to the original model. The coefficients in the oversampling and SMOTE models are higher for all features compared to the original model, with SMOTE showing the highest coefficient values. This difference suggests that the relationship between the risk factors and depression are stronger for balanced classes, especially when SMOTE is used. Thus, SMOTE appears to improve the prediction of depression compared to the original model by effectively handling class imbalance in the logistic regression model.

7. Conclusion

The study identifies pressure and age as the most influential predictors of depression, while other important factors include financial stress, employment/student status, and suicidal thoughts. Higher levels of pressure, younger age, increased financial stress, student status, and more frequent suicidal thoughts all contribute to a higher risk of depression. This finding emphasizes the importance of considering multiple risk factors when developing personalized treatment strategies, rather than focusing on a single, typical profile. It also helps to identify vulnerable groups that need special attention and targeted intervention strategies, such as young students under pressure or those who are financially unable to support themselves. Thus, understanding the correlation between potential risk factors and depression is essential for developing personalized and effective treatment plans to prevent recurrence or chronic depression. A predictive model is needed to assess how specific risk factors influence the risk of depression, both individually and in combination, so that society can focus on these factors and allocate resources more effectively to reduce the prevalence of depression. Early identification of individuals at higher risk also promotes early and targeted intervention, increasing its effectiveness. The predictive model developed in the study has a high accuracy of 91%. By predicting depression risk before symptoms such as somatization are fully manifest, this model may lead to earlier and more proactive interventions, reducing the overall impact of depression on individuals and those around them. While the model is accurate and has potential, improving its precision can increase the accuracy of identifying individuals who are truly at risk for depression, reducing the need for unnecessary interventions.

Future research should focus on further breaking down risk factors into smaller, more specific categories to better understand how each one influences the depression risk so that early interventions can target even more specific factors or groups. For example, pressure can be broken down into work/study pressure, family pressure, and social comparison pressure; age can be broken down into children, teenagers, young adults, adults, and seniors; financial stress can be analyzed in terms of expected versus actual income and necessary versus discretionary expenses; and employment/student status can be examined based on jobs/grade levels, and type of school/company. Longitudinal studies are also needed to track how these risk factors evolve over time and their long-term impact on recurrence or chronic depression. In addition, examining other potential risk factors, such as

DOI: 10.54254/2755-2721/157/2025.PO23118

personality traits, stimulant consumption, and relationships, can provide a more complete understanding of an individual's life and its influence on the depression risk. To ensure a broad application of the model, it should be tested with data from different regions, considering cultural, ethnic, and environmental differences that may influence the depression risk. Necessary modifications should be made to better adapt the model, and the effectiveness of the early intervention strategies and personalized treatment developed in this study should be tested multiple times to confirm their impact. Finally, previous studies have shown a relatively high overlap in populations with symptoms of depression, anxiety, and stress, highlighting similarities in their risk factors and symptoms.[8] This finding suggests that future models should examine the relationships among depression, anxiety, and stress and assess the importance of risk factors for each disorder, both individually and in combination.

References

[1] Reinert, Maddy, et al. "The State of Mental Health in America 2024." Mental Health America, 1 Jul. 2024, http://hdl.handle.net/10713/22688.

[2] Zhukovsky, Peter, et al. "Generalizability of Treatment Outcome Prediction Across Antidepressant Treatment Tria ls in Depression." JAMA Network Open, vol. 8, no. 3, 20 Mar. 2025, JAMA Network, https://doi.org/10.1001/jama networkopen.2025.1310.

[3] Stringaris, Argyris. "Editorial: What is depression?" The Journal of Child Psychology and Psychiatry, vol. 58, n o. 12, 17 Nov. 2017, pp. 1287–1289, The Association for Child and Adolescent Mental Health, https://doi.org/10.1 111/jcpp.12844.

[4] Olfson, Mark, et al. "Treatment of Adult Depression in the United States." JAMA Internal Medicine, vol. 176, no. 10, 29 Aug. 2016, pp. 1482–1491, JAMA Network, https://doi.org/10.1001/jamainternmed.2016.5057.

[5] Josefsson, T., et al. "Physical exercise intervention in depressive disorders: Meta-analysis and systematic review." Scandinavian Journal of Medicine & Science in Sports, vol. 24, no. 2, 30 Jan. 2013, John Wiley & Sons, Inc., https://doi.org/10.1111/sms.12050.

[6] Mofatteh, Mohammad. "Risk factors associated with stress, anxiety, and depression among university undergraduate students." AIMS public health, vol. 8, no. 1, 25 Dec. 2020, pp. 36–65, National Library of Medicine, https://doi.org/10.3934/publichealth.2021004.

[7] Hammen, Constance. "Risk Factors for Depression: An Autobiographical Review." Annual Review of Clinical Ps ychology, vol. 14, 12 Jan. 2018, pp. 1–28, Annual Reviews, https://doi.org/10.1146/annurev-clinpsy-050817-0848 11.

[8] Ramón-Arbués, Enrique, et al. "The Prevalence of Depression, Anxiety and Stress and Their Associated Factors i n College Students." International Journal of Environmental Research and Public Health, vol. 17, no. 19, 24 Sep. 2020, MDPI, https://doi.org/10.3390/ijerph17197001.

[9] Wang, Jinghui, et al. "Prevalence of depression and depressive symptoms among outpatients: a systematic review and meta-analysis." BMJ Open, vol. 7, no. 8, 5 July 2017, BMJ Publishing Group Ltd., https://doi.org/10.1136/b mjopen-2017-017173.

[10] McGorry, Patrick D., and Cristina Mei. "Early intervention in youth mental health: progress and future direction s." BMJ Mental Health, vol. 21, no. 4, 1 Nov. 2018, pp. 182–184, BMJ Publishing Group Ltd., https://doi.org/10.1 136/ebmental-2018-300060.

[11] Ormel, Johan, et al. "More treatment but no less depression: The treatment-prevalence paradox." Clinical Psychology Review, vol. 91, 24 Dec. 2021, ScienceDirect, https://doi.org/10.1016/j.cpr.2021.102111.

[12] Fazel, Mina, et al. "Mental health interventions in schools in high-income countries." The Lancet Psychiatry, vol. 1, no. 5, pp. 377–387, https://doi.org/10.1016/S2215-0366(14)70312-8.