Predicting Depression

Simon Rupp, YongJae Lee, Aryavrat Gupta

December 18, 2024

Abstract

Depression is a pervasive mental health issue that plagues the lives of millions worldwide. According to the Depression and Bipolar Support Alliance (DBSA), an estimated 21 million (8.3%) adults in the United States have had at least one major depressive episode[1]. Using data from the National Health and Nutrition Examination Survey (NHANES)[2], our project aims to develop predictive models for depression symptoms and identify the most significant features that contribute to their prediction. We implemented models including Random Forest, Logistic Regression, SVM, XGBoost, LightGBM, and a Stacking Classifier, with cross-validation used for evaluation. Feature importance analysis was conducted using both Gini importance and permutation importance methods. Our findings highlight significant health, lifestyle, and socioeconomic predictors, offering insights that can guide targeted interventions and preventive measures.

1 Introduction

Depression (major depressive disorder), often described in overly simplistic or ambiguous terms, is far more complex and impactful than many definitions convey. As a common and serious mental health disorder, major depressive disorder profoundly influences how individuals feel, think, act, and perceive the world around them[3]. Beyond its clinical definition, depression represents a substantial global health challenge, contributing to diminished quality of life, heightened healthcare costs, and significant societal burden.

Current approaches to depression diagnosis and prediction face significant challenges, including the overlap of symptoms with other disorders, variability in presentation across individuals, the absence of objective biomarkers, and the influence of social stigma and cultural factors[4]. These challenges are compounded by the reliance on subjective self-reporting, which, while imperfect, remains one of the most accessible and widely used methods for collecting mental health data. Traditional predictive models often focus on identifying patterns and correlations, which can limit their practical utility in designing effective interventions.

Our project seeks to address these challenges by leveraging self-reported questionnaire data, along with demographic, socioeconomic, dietary, and lifestyle information, from the NHANES dataset. While we acknowledge the limitations of self-reported data, its breadth allows us to capture diverse factors influencing depression. By identifying key predictors of depression symptoms, we aim to enhance the understanding of depression risk factors and support the development of targeted treatment and preventive strategies, addressing a critical gap in mental health research.

2 Problem Statement

The primary goal of this project is twofold:

1. Develop predictive models for depression symptoms using NHANES data, incorporating features such as nutrient intake, physical activity, sleep, and demographic information.

2. Identify the most significant features associated with depression outcomes to inform actionable insights.

Depression is a complex condition influenced by multiple variables[5]. Identifying key predictors is crucial for understanding depression risk factors and developing effective treatment and prevention strategies.

3 Significance

This project is a small step towards uncovering the factors associated with depressive symptoms, and its findings highlight the need for further research to fully realize their implications.

1. **Clinical Implications for Treatment Strategies** - This study identifies potential predictors of depression, offering preliminary insights that could guide future clinical research. While these findings are limited, they pave the way for further exploration into how key factors such as diet inform personalized treatment approaches and early intervention efforts.

2. **Public Health Policy Implications** - By shedding light on the association between socioeconomic, lifestyle, and demographic variables with depression, this project suggests areas where public health initiatives could focus their attention. Further research is crucial to translate these insights into actionable, evidence-based community programs that address mental health disparities.

3. **Potential for Preventive Intervention** - The results emphasize the importance of studying modifiable factors, such as lifestyle choices and physical activity, in the context of depression. Continued investigation into how lifestyle adjustments might contribute to prevention strategies offers a hopeful direction for future mental health research.

4 Methodology

4.1 Data Description

We used publicly available data from NHANES, which can be accessed through the CDC NHANES website. This dataset includes comprehensive health, lifestyle, and demographic variables. Depression outcomes are assessed using the PHQ-9 questionnaire, providing a standardized measure of depression severity.

The primary dataset that determines the target values for this project is the DPQ_L.xpt file, which contains depression-related questions from the PHQ-9 assessment. This nine-item depression screening questionnaire[6], referred to as the Patient Health Questionnaire (PHQ-9), evaluates the frequency of depressive symptoms experienced by participants over the past two weeks. Each symptom is rated on a scale of 0 to 3, corresponding to the response categories: “not at all,” “several days,” “more than half the days,” and ”nearly every day.” An additional functional impairment question is asked if any symptom is endorsed. The PHQ-9 incorporates DSM-IV diagnostic criteria for depression[7], making it a widely used tool for assessing depressive symptoms. In this study, the PHQ-9 scores are used to determine the depression status of participants, serving as the target variable (outcome).

4.1.1 Combining Datasets

To enhance the predictive capabilities of our model, data from various NHANES subsets were integrated. These subsets include:

- **Demographic Data:** Information about participants, including household and individual demographic details:
  - Gender, age, race/Hispanic origin, education, marital status, and military service status.
  - Household details such as pregnancy status, total number of household members, and the household reference person’s demographic information.
  - Economic indicators, such as the ratio of family income to poverty guidelines.
  - Interview and examination sample weights, masked variance units, and the examination time period.
- **Dietary Data:** Data from 24-hour dietary recall sessions, estimating types and amounts of foods and beverages consumed, as well as associated nutrient and energy intakes. Additional questions include:

  - Salt usage and dietary habits relative to usual intake.
  - Frequency of fish and shellfish consumption over the past 30 days.
  - Special diet information and dietary habits for participants aged 1 year or older, including the use of proxies for young children.

- **Examination Data:** Anthropometric measurements, including:
  - Weight (all ages), height (2 years and older), and recumbent length (birth through 47 months).
  - Circumferences (waist, hip, mid-upper arm) and limb lengths (upper arm, upper leg) for relevant age groups.
  - Head circumference for infants (birth through 6 months).

- **Questionnaire Data:** Information related to lifestyle and behavioral factors, including sleep patterns, smoking and alcohol consumption, and physical activity.

All datasets were merged using the unique SEQN identifier assigned to each participant. The resulting consolidated dataframe combines a diverse array of health, dietary, demographic, and behavioral data, providing a comprehensive foundation for predicting depressive symptoms.

4.2 Preprocessing

The NHANES dataset underwent a series of preprocessing steps to handle missing values, standardize features, and ensure the data was suitable for machine learning models. These steps included the following:

4.2.1 Handling Missing Data

During the integration process, a missing data filtering technique was applied, where columns with a missingness percentage greater than the defined missingness-threshold constant were removed to ensure data quality and minimize potential biases. This thresholding assumes that missing data is not completely random, and removing highly incomplete columns prevents imputations from introducing undue uncertainty.

We consider the data to follow a Missing at Random (MAR) model, as the likelihood of missing data depends on observed variables rather than unobserved factors. For example, in the NHANES dataset, missing data in dietary intake or physical activity features might correlate with demographic or socioeconomic factors like age, income, or education level, which are observed (refer to Figure 1). This assumption allows the remaining observed data to be used for reliable imputation and analysis.

4.2.2 Imputation of Missing Values

For remaining columns, we tested imputing missing values using a variety of methods. The chosen strategy was mean imputation. This was chosen based on model performance. We tested mean, median, and most frequent imputation as well as K-Nearest Neighbors (KNN) imputation. Empirically, mean imputation was the best, so it is what we ultimately went with.

4.2.3 Encoding and Standardizing Features

Non-numeric columns, such as those in the sleep questionnaire, were converted to numeric format by extracting time values and handling non-standard entries. Continuous features were standardized to normalize ranges and improve model performance.

Similarly, to ensure consistency in physical activity features, units of frequency (daily, weekly, monthly, yearly) were converted to a common annualized format using predefined conversion factors. For example, a participant's reported weekly activity was multiplied by 52.14 to estimate annual frequency.

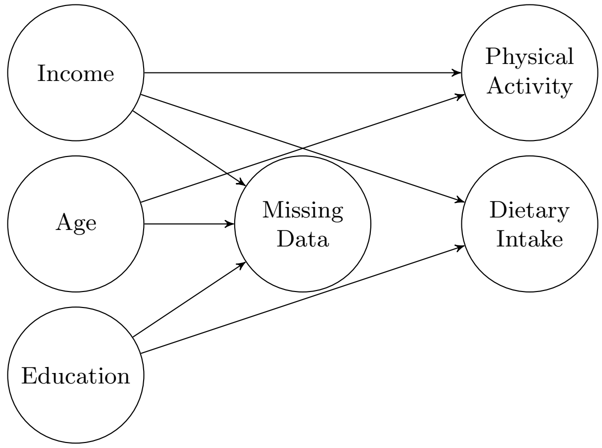

Figure 1: Example DAG Representing a MAR Model in NHANES Data. Observed variables such as income, age, and education might influence both dietary intake, physical activity, and the likelihood of missing data.

4.2.4 Target Variable Preparation

Depression was quantified using the PHQ-9 total score (called "DPQSUM" in the code), which served as the target variable. All individual PHQ-9 question columns were excluded since they directly determined the total score. The target variable was split into two formats:

- **Binary Classification:** Scores of 0-9 were classified as "not depressed," while scores of 10-27 were classified as "depressed."
- **Continuous Classification:** Scores were divided into five severity levels: minimal (0-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20-27).

4.2.5 Balancing the Dataset

As shown in Figure 2, the distribution of PHQ-9 total scores is heavily skewed toward lower values, indicating that the majority of participants were classified as not depressed (scores between 0 and 9). This significant class imbalance necessitated balancing techniques to ensure fair model training and evaluation.

Before balancing, the dataset exhibited a significant class imbalance in both the binary and continuous target distributions. Table 1 shows the original distributions. To address the class imbalance in the binary target, the majority class (not depressed) was downsampled to match the minority class (depressed). This resulted in a balanced distribution, as shown in Table 2.

Table 1: Original Target Distributions

Target	Class/Level	Count
Binary (DEPRESSED)	0 (Not Depressed)	4099
	1 (Depressed)	672
Continuous (DEPRESSION_LEVEL)	1 (Minimal)	3235
	2 (Mild)	864
	3 (Moderate)	411
	4 (Moderately Severe)	159
	5 (Severe)	102

While the real-world prevalence of depression is lower than 50%, a balanced dataset (50-50 split) was created to prevent model bias and ensure the model can accurately learn to predict both "depressed" and "not depressed" outcomes.

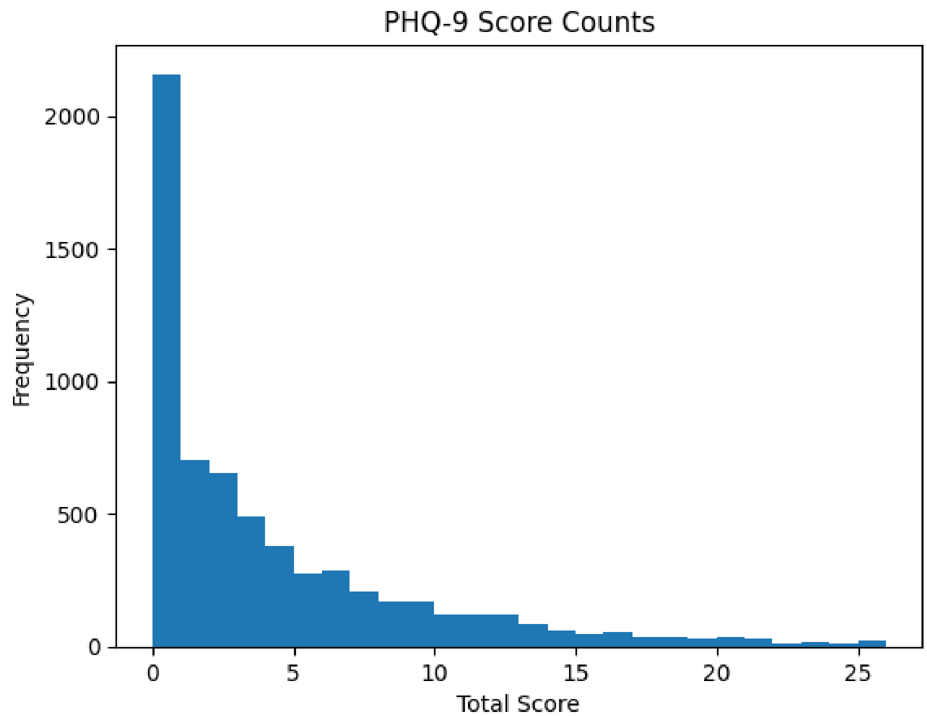

Figure 2: Distribution of PHQ-9 Total Scores. A significant portion of participants have low scores, indicating they are not depressed.

Table 2: New Binary Target Distribution After Balancing

Class	**Count**
0 (Not Depressed)	672
1 (Depressed)	672

4.2.6 Train-Test Splitting

The processed data was split into training and testing sets (80:20 split) to evaluate model performance. Predictor variables (X) included all features except the target columns (DEPRESSED and DEPRESSION_LEVEL), which served as the binary and continuous targets, respectively.

This preprocessing workflow ensured the dataset was well-structured, balanced, and ready for machine learning modeling. It addressed common challenges in handling real-world datasets, including missing data, feature inconsistencies, and class imbalances.

4.3 Model Selection

To predict depression symptoms, we evaluated several machine learning models: Random Forest, Logistic Regression, Support Vector Machines (SVM), XGBoost, LightGBM, and a Stacking Classifier. Each model was chosen based on its ability to handle tabular data, scalability, and interpretability. Cross-validation (5-fold) was used to evaluate model performance and ensure generalizability.

4.3.1 Random Forest

Random Forest is an ensemble method that builds multiple decision trees and aggregates their results to improve accuracy and reduce overfitting. Given its robustness to noise and feature importance capabilities, it was well-suited for our dataset. Mathematically, the prediction for a classification task is:

$$\hat{y} = \text{argmax}_c \left(\frac{1}{T} \sum_{t=1}^{T} I(h_t(x) = c) \right),$$

where $h_t(x)$ is the output of tree t, T is the total number of trees, c is the class label, and I is the indicator function.

For our case, we used $T = 100$ trees with a maximum depth of 3 to balance performance and interpretability.

4.3.2 Logistic Regression

Logistic Regression is a linear model used for binary classification. It models the probability that a sample belongs to a particular class using the logistic (sigmoid) function:

$$P(y = 1 \mid x) = \frac{1}{1 + \exp(-w^T x - b)},$$

where w represents the weight vector, x is the input feature vector, and b is the bias term.

Logistic Regression was chosen as a baseline model due to its simplicity, interpretability, and effectiveness in linearly separable datasets. As mentioned before, standardization was applied to the features before training.

4.3.3 Support Vector Machines (SVM)

SVMs aim to find a hyperplane that maximizes the margin between two classes. For non-linearly separable data, a radial basis function (RBF) kernel was applied, transforming the data into a higher-dimensional space. Mathematically, the optimization problem for SVM is:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \quad \text{s.t.} \quad y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0,$$

where $\phi(x)$ is the kernel function, C is a regularization parameter, and ξ_i are slack variables to handle misclassification.

SVM was selected for its ability to handle complex, non-linear relationships.

4.3.4 XGBoost and LightGBM

XGBoost and LightGBM are gradient-boosting frameworks that build an ensemble of weak decision trees sequentially. At each iteration, the model minimizes a loss function, such as the binary cross-entropy loss for classification:

$$L = -\sum_{i=1}^{n} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right],$$

where y_i is the true label, and $\hat{y}_i$ is the predicted probability.

LightGBM was specifically chosen for its efficiency on large datasets.

4.3.5 Stacking Classifier

The Stacking Classifier combines multiple base models (in our case: Random Forest, XGBoost, and LightGBM) and uses a meta-model (we chose: Logistic Regression) to aggregate their predictions. This ensemble approach leverages the strengths of each base learner to improve performance. The meta-model learns from the base models' outputs to make the final prediction.

Let $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m$ be the predictions of m base models. The final prediction is given as:

$$\hat{y} = g(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m),$$

where g is the meta-model (Logistic Regression in our case).

Stacking was chosen for its ability to combine diverse models and enhance overall predictive performance.

4.3.6 Model Evaluation

Cross-validation (5-fold) was performed for each model, and the mean accuracy score was reported. The results are summarized in Table 3.

Table 3: Model Performance Results (Mean Accuracy Scores)

Model	Mean Accuracy Score
Random Forest	0.7366
Logistic Regression	0.7158
Support Vector Machine (SVM)	0.7135
XGBoost	0.7433
LightGBM	0.7403
Stacked Model	0.7441

4.4 Experimental Setup

We conducted experiments to predict depression status and severity based on features from the NHANES dataset. The experiments were performed in two phases:

1. **Binary Classification:** Predict the presence of depression (yes/no) based on the PHQ-9 total score, where scores ≥ 10 indicated depression.
2. **Feature Importance Analysis:** Identify the most significant features contributing to the predictions using Gini importance and permutation importance.

We tested several machine learning models, including Random Forest, Logistic Regression, Support Vector Machine (SVM), XGBoost, LightGBM, and a Stacking Classifier. A 5-fold cross-validation strategy was applied to ensure robust evaluation of each model.

The following code snippet demonstrates the general process for determining cross-validation scores:

```
from sklearn.model_selection import cross_val_score

# Example for LightGBM
from lightgbm import LGBMClassifier

model = LGBMClassifier(random_state=42)
cv_scores = cross_val_score(model, X, y_binary, cv=5, scoring='accuracy')

print("Mean Accuracy Score:", cv_scores.mean())
```

The above process was repeated for all models to compute their respective mean accuracy scores across the folds.

5 Results

The results of our experiments are summarized in Table 3. We tested multiple machine learning models to predict depression presence (binary classification) and identify significant features.

5.1 Model Performance

Among the models tested, the **Stacking Classifier** achieved the highest mean accuracy of **74.41%**, outperforming all other models. This performance can be attributed to its ability to combine the strengths of multiple base models (Random Forest, XGBoost, and LightGBM) while reducing individual model biases and better managing feature complexity.

Key Observations:

- **Stacking Classifier (74.41%):**
  - Leverages the diversity and complementary strengths of multiple models.
  - Reduces biases inherent in standalone models.
  - Better handles non-linear relationships and complex feature interactions.
- **XGBoost (74.33%):** As the second-best performer, XGBoost demonstrated the advantages of gradient boosting:
  - Excellent handling of non-linear relationships.
  - Robust to outliers and missing values.
  - Effectively captures complex feature interactions.
- **LightGBM (74.03%):** LightGBM's performance was close to XGBoost, showcasing its efficiency in handling large-scale tabular data with categorical features.
- **Random Forest (73.66%):** The Random Forest model performed well due to its robustness and ability to handle feature importance, though it lacked the boosting mechanisms of XGBoost and LightGBM.

  As shown in Table 4, the Random Forest model demonstrated balanced performance across both classes. For Class 0 (not depressed), the model achieved a precision of 0.79, recall of 0.76, and an F1-score of 0.78, indicating slightly better performance in correctly predicting non-depressed individuals, though some true positives were missed due to lower recall. For Class 1 (depressed), the precision was 0.74, recall was 0.77, and the F1-score was 0.75, reflecting balanced performance and high recall, which highlights the model's ability to detect most cases of depression. The macro and weighted averages for precision, recall, and F1-score were consistent at 0.77, suggesting that the model performed uniformly across both classes without significant bias toward one class. Additionally, the use of a balanced dataset, with equal samples for depressed and non-depressed cases, ensured that the model was not skewed toward the majority class, resulting in reliable predictions for both groups.

Table 4: Binary Target Classification Report for Random Forest

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0 (Not Depressed)	0.79	0.76	0.78	144
1 (Depressed)	0.74	0.77	0.75	125
Accuracy		0.77		**269**
Macro Avg	0.76	0.77	0.77	
Weighted Avg	0.77	0.77	0.77	

- **Logistic Regression (71.58%) and SVM (71.35%):** These models underperformed due to:
  - Inability to fully capture complex non-linear relationships in the dataset.
  - High feature dimensionality, which limited their predictive capacity.
  - Possible need for additional feature engineering and transformation.

Overall, the results indicate that ensemble models like the Stacking Classifier and boosting-based approaches (XGBoost and LightGBM) are more effective for handling the complexity and non-linearity of the NHANES dataset features.

5.2 Feature Importance Analysis

To interpret the predictions and better understand the factors associated with depression, we conducted feature importance analysis using the Random Forest model. Two methods were applied:

- **Gini Importance:** Measures how much each feature reduces node impurity across all trees in the Random Forest.
- **Permutation Importance:** Measures the impact of each feature on model accuracy by shuffling feature values and observing the change in performance.

5.2.1 Top 20 Features Based on Gini Importance

Figure 3 presents the top 20 features identified using Gini importance. These features provide insights into the factors most strongly associated with depression predictions.

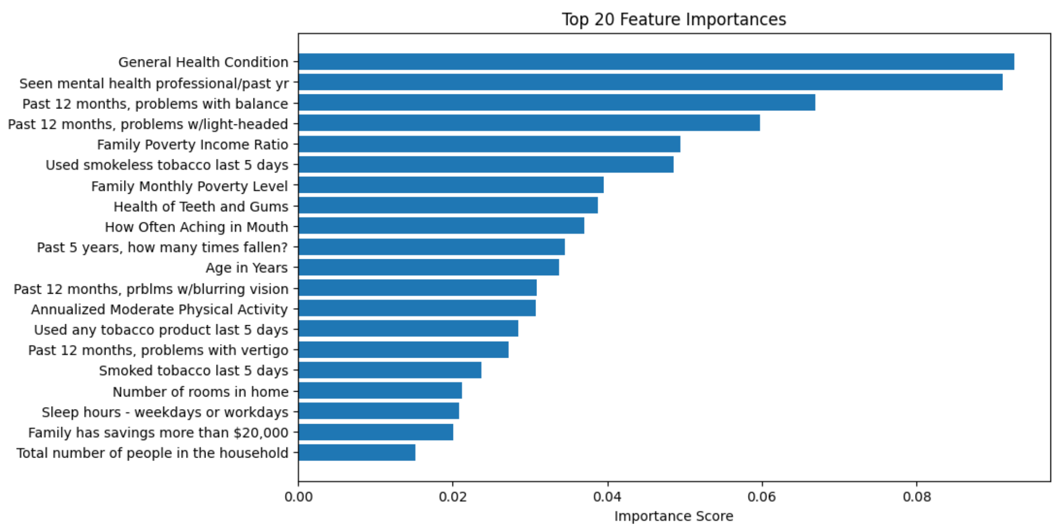

Figure 3: Top 20 Feature Importances Based on Gini Importance. General health condition and mental health professional visits emerged as the most significant predictors.

Key Observations:

- **General Health Condition:** The strongest predictor, indicating a strong association between overall health perception and depression.
- **Recent Mental Health Professional Visits:** Access to mental health services emerged as a highly significant feature.
- **Socioeconomic Status:** Family poverty income ratio and monthly poverty level highlight the impact of financial stress on mental health.
- **Lifestyle Factors:** Physical activity and sleep duration further emphasize the role of modifiable behaviors in depression risk.

Unusual and Interesting Insights: Interestingly, features such as **health of teeth and gums**, **aching in the mouth**, and **recent tobacco use** emerged as important predictors. While at first glance these observations may seem unusual, they could indicate indirect relationships:

- Poor dental health or frequent mouth pain may reflect larger systemic health issues or stressors, which are often associated with depression.
- Tobacco use may serve as a coping mechanism for individuals experiencing depressive symptoms, further connecting behavioral habits to mental health.

- The presence of these features emphasizes the importance of examining subtle, seemingly unrelated health factors in mental health predictions.

However, this analysis of strange findings highlights a tendency to perceive patterns where none exist. It is possible that these features do not directly impact depression and appeared as important predictors due to noise in the data or correlations with other more meaningful variables. This highlights the need for further investigation and validation to distinguish true relationships from spurious associations.

5.2.2 Top 10 Features Based on Permutation Importance

Permutation importance (Figure 4) provides a complementary perspective by measuring the impact of individual features on model accuracy. This analysis reinforces the findings from Gini importance while offering additional insights.

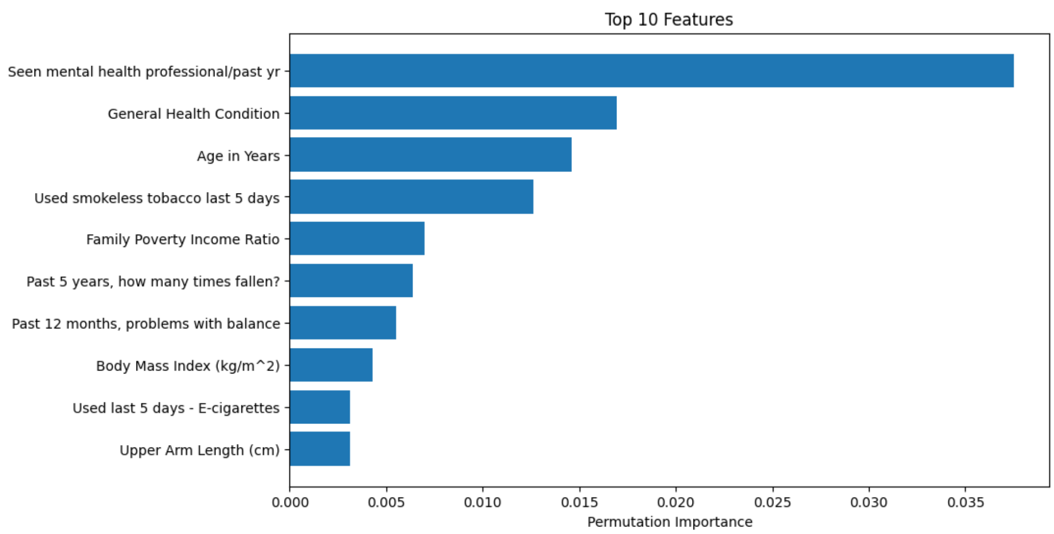

Figure 4: Top 10 Feature Importances Based on Permutation Importance. Seen mental health professional visits and general health condition were the top predictors.

Key Insights:

- **Seen Mental Health Professional (Past Year):** This feature had the highest impact on model performance, highlighting the importance of mental health service utilization.
- **General Health Condition:** Consistently ranked among the top predictors across both importance methods.
- **Age and Socioeconomic Factors:** Features such as age, family poverty income ratio, and physical health markers (e.g., BMI, tobacco use) contributed significantly to model predictions.

Similar to our earlier analysis of the unusual prominence of the top 20 features, attributes such as **frequency of falling** and **BMI** surfaced among the top 10 features based on permutation importance. However, these features seemingly lack predictive relevance for diagnosing depression. This discrepancy could be attributed to the presence of confounding variables or noise in the dataset, which may inadvertently influence the model's learning process, leading to overfitting or misleading interpretations of feature importance.

5.3 Summary of Findings

The feature importance analysis underscores the multifaceted nature of depression, with key predictors spanning health perception, socioeconomic status, and lifestyle behaviors. The consistency of top features across both Gini and permutation importance methods enhances confidence in the robustness of these findings. These insights highlight actionable areas for targeted mental health interventions, such as improving access to mental health care and addressing modifiable lifestyle factors. Additionally, the identification of unexpected features like frequency of falling and BMI as important predictors warrants further investigation to understand their relationship with depression, suggesting that there may be underlying factors or interactions within the data that are not immediately apparent.

6 Challenges

The project faced several challenges:

- **Data Imbalance:** The dataset contained fewer samples in higher depression severity categories. To avoid a predictor that always classified patients as not depressed, we had to remove many not depressed patients to balance out the classes. This limited the amount of information we could use to train our models.
- **Feature Complexity:** Interpreting relationships among highly correlated variables required careful analysis.
- **Missing Data:** There was a high degree of missingness in the NHANES dataset. Many patients seemed to decline to answer or just didn't have a values associated with certain metrics. This forced us to remove many features with high missingness and impute missing values. Removing features and imputing values certainly limited our ability to classify patients correctly.
- **Self-Reported Data:** Many features relied on self-reported inputs, potentially introducing reporting bias or inaccuracies.
- **Model Interpretability:** Complex models like Stacking Classifiers and XGBoost posed challenges in interpreting the impact of individual features.
- **Continuous Classification:** Earlier we mentioned that we created the target variables in two ways: through a binary (depressed or not depressed) classification as well as a more continuous (minimal, mild, moderate, moderately severe, or severe) classification. Given the somewhat average prediction results from binary classification ($\approx 74\%$), we accepted that creating a model to predict the continuous targets (1-5) would be unsuccessful.

7 Implications and Future Work

This project highlights the importance of leveraging predictive models to better understand depression risk and guide mental health interventions. While our approach focused on identifying significant predictors, further work is needed to refine these insights and translate them into actionable strategies. The key implications and directions for future work are as follows:

Clinical Implications:

- **Applications in Clinical Settings:** Predictive models like those developed in this project can assist clinicians in identifying at-risk individuals and prioritizing early interventions[8].
- **Prevention Strategies:** Insights into modifiable factors such as physical activity, sleep duration, and socioeconomic status can inform public health campaigns aimed at reducing depression prevalence[9].
- **Treatment Optimization:** Personalized treatment plans can be developed by integrating predictive insights into clinical decision-making, targeting the most influential factors for each individual[10].

Future Work:

- **Extended Feature Engineering:** Incorporate additional features such as genetic, environmental, and social determinants of health to improve model accuracy and relevance[11].

- **Advanced Analytical Methods:** Explore more sophisticated causal analysis techniques to strengthen the differentiation between causation and correlation[12].
- **Longitudinal Data Analysis:** Use time-series data to investigate how predictors and outcomes evolve over time, offering deeper insights into depression progression[13].
- **Clinical Validation Studies:** Validate the findings through collaboration with healthcare practitioners and integration with real-world datasets.
- **Healthcare System Integration:** Develop scalable solutions to integrate these predictive models into existing healthcare infrastructures, enabling widespread adoption.

8 Conclusion

This project demonstrated the potential of machine learning models, particularly ensemble methods like the Stacking Classifier, in predicting depression and identifying key predictors. By analyzing features related to health, lifestyle, and socioeconomic factors, our findings provide actionable insights for both clinical practice and public health strategies.

The work underscores the complexity of depression, driven by interdependent variables that span biological, psychological, and environmental domains. While our models achieved a reasonable accuracy of 74.41%, this highlights the challenge of addressing mental health with limited data and underscores the need for further exploration. Key contributions: i) Identified significant predictors such as general health condition, socioeconomic factors, and lifestyle habits, ii) Demonstrated the effectiveness of ensemble methods in managing complex and non-linear data relationship, iii) Highlighted the importance of balancing datasets to improve predictive reliability across different populations.

This study contributes to the growing intersection of machine learning and mental health research, emphasizing the potential for predictive models to improve understanding and treatment of depression[14]. Future efforts should focus on expanding datasets, validating findings in clinical settings, and integrating models into practical applications, ultimately bridging the gap between research and real-world impact.

References

[1] Depression and Bipolar Support Alliance. Depression statistics. https://www.dbsalliance.org/education/depression/statistics/, 2024.

[2] Centers for Disease Control and Prevention. Nhanes august 2021-august 2023. https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?Cycle=2021-2023, 2021-2023.

[3] American Psychiatric Association. What is depression? https://www.psychiatry.org/patients-families/depression/what-is-depression, 2024.

[4] X. Liu and K. Jiang. Why is diagnosing mdd challenging? *Shanghai Archives of Psychiatry*, 28(6):343–345, December 25 2016.

[5] J. M. Llorente, B. Oliván-Blázquez, M. Zuñiga-Antón, B. Masluk, E. Andrés, J. García-Campayo, and R. Magallón-Botaya. Variability of the prevalence of depression in function of sociodemographic and environmental factors: Ecological model. *Frontiers in Psychology*, 9:2182, November 12 2018.

[6] K. Kroenke, R. L. Spitzer, and J. B. Williams. The phq-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9):606–613, September 2001.

[7] R. L. Spitzer, K. Kroenke, and J. B. Williams. Validation and utility of a self-report version of prime-md: the phq primary care study. primary care evaluation of mental disorders. patient health questionnaire. *JAMA*, 282(18):1737–1744, November 10 1999.

[8] S. Eloranta and M. Boman. Predictive models for clinical decision making: Deep dives in practical machine learning. *Journal of Internal Medicine*, 292(2):278–295, August 2022.

[9] A. R. Ramos, A. G. Wheaton, and D. A. Johnson. Sleep deprivation, sleep disorders, and chronic disease. *Preventing Chronic Disease*, 20:E77, August 31 2023.

[10] S. A. Alowais, S. S. Alghamdi, N. Alsuhebany, T. Alqahtani, A. I. Alshaya, S. N. Almohareb, A. Aldairem, M. Alrashed, K. Bin Saleh, H. A. Badreldin, M. S. Al Yami, S. Al Harbi, and A. M. Albekairy. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education*, 23(1):689, September 22 2023.

[11] E. E. Lee, J. Torous, M. De Choudhury, C. A. Depp, S. A. Graham, H. C. Kim, M. P. Paulus, J. H. Krystal, and D. V. Jeste. Artificial intelligence for mental health care: Clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(9):856–864, September 2021.

[12] K. S. Kendler and C. O. Gardner. Dependent stressful life events and prior depressive episodes in the prediction of major depression: the problem of causal inference in psychiatric epidemiology. *Archives of General Psychiatry*, 67(11):1120–1127, November 2010.

[13] D. C. R. van Zelst, E. M. Veltman, D. Rhebergen, P. Naarding, A. A. L. Kok, N. R. Ottenheim, and E. J. Giltay. Network structure of time-varying depressive symptoms through dynamic time warp analysis in late-life depression. *International Journal of Geriatric Psychiatry*, 37(9):e5787, September 2022.

[14] N. K. Iyortsuun, S. H. Kim, M. Jhon, H. J. Yang, and S. Pant. A review of machine learning and deep learning approaches on mental health diagnosis. *Healthcare (Basel)*, 11(3):285, January 17 2023.