Optimise Diagnosis of Depression based on Hypertension and Sleep Apnea Factors

Ann Maria John*, Madhulika Dutta†, Priyanka Bhyregowda‡, Shruti Badrinarayanan§
Department of Applied Data Science, San Jose State University*†‡§
annmaria.john@sjsu.edu*, madhulika.dutta@sjsu.edu†, priyanka.bhyregowda@sjsu.edu‡, shruti.badrinarayanan@sjsu.edu§

***Abstract*—Analyzing sleep apnea and hypertension as indicators of depression risk is significant owing to their prevalent coexistence with depression. Determining the presence of these disorders as risk factors for depression may facilitate early intervention, improving mental health outcomes and lessening the impact of depression. Better treatment outcomes can be achieved by addressing the interconnected physical and mental health problems using this approach, which is in line with holistic healthcare. The project's overall goal is to advance understanding of these relationships, leading to improved diagnostic accuracy and optimized interventions for affected individuals.**

***Index Terms*—Exploratory Data Analysis, Population, NHANES, Obstructive Sleep Apnea, Hypertension, Depression**

I. Introduction

Depression is a prevalent mental health condition with serious repercussions for sufferers and society [1],[2]. It is characterized by symptoms such as chronic fatigue, poor mood, and decreased interest [1]. It is a primary cause of disability, accounting for around 1 million deaths yearly, and affects over 300 million people worldwide [1],[3] making up 4.4% of the total population [1]. Despite its significant impact, a significant number of people, especially in low- and middle-income countries continue to go undiagnosed and untreated [1]. Elderly and middle-aged people are disproportionately affected as well. Negative effects of depression include lower quality of life [2],[3],[4], increased risks for mortality and dementia [3], raising healthcare costs [2], adherence to treatments and health behavior [4]. To lessen the disorder's negative effects, early detection and prevention are essential [1],[4]. The US alone spends up to $210 billion a year on treating depression, which has significant social repercussions [1].

Excessive daytime sleepiness (EDS), irregular sleep patterns, short sleep duration and delayed sleep phase, are closely associated with depression [2]. Depression is also associated with hypertension [4],[5]. The duration and severity of hypertension, coupled with hospitalization history, are predictive factors for the higher prevalence of depression that is linked to the former [6].

The project aims to optimize the diagnosis of depression based on hypertension and sleep apnea as factors both of which are known to coexist with depression. Firstly, it raises awareness about how these health factors can impact our mental well-being [1]. When people understand the potential links between these conditions and depression, they're more likely to take proactive steps to care for their health.

Secondly, the insights gained from this project can guide hospitals in offering more comprehensive care. Healthcare professionals can reduce the incidence of depression in people with sleep apnea or hypertension by identifying these illnesses as potential risk factors for depression and using early screening and management strategies [4]. This proactive approach could lessen the impact of depression on people and society, at the same time improving mental health outcomes.

Overall, the motivation for the project lies in the potential to advance our understanding of the intricate relationships between depression and factors like sleep apnea and hypertension, resulting in improved diagnostic accuracy, early intervention, and better mental health outcomes for those who are impacted by them.

II. Related Work

Research suggests that there is evidence that a dose-response association exists between the severity of Obstructive Sleep Apnea Hypopnea Syndrome (OSAHS) and the risk of depression [7]. [8] explores the connection between obstructive sleep apnea (OSA) and hypertension. This work finds plausible biological reasons for hypertension in OSA patients and notes modest blood pressure reductions after OSA treatment. While larger prospective trials are needed to firmly establish OSA as a direct cause of hypertension, the potential benefits of treating OSA are clear.

Studies have underscored the complex interrelationships between sleep disorders and mental health, particularly highlighting the associations between insomnia, obstructive sleep apnea (OSA), and depression. Building on prior research, a large-scale analysis using the National Health and Nutrition Examination Survey (NHANES) data (2005-2008) was conducted, investigating the link between sleep disorders and depression in a representative sample of US adults [8]. This study examined the independent and combined effects of insomnia and OSA on depression risk, while controlling for potential confounders like demographics, lifestyle habits (smoking), and body mass index (BMI). The results demonstrated a significant association between both insomnia and OSA, alone or co-occurring, with increased depression prevalence.

To address the challenge of predicting depression in hypertensive populations, machine learning (ML) techniques were leveraged [9]. Data from the National Health and Nutrition Examination Survey (NHANES, 2011-2020) including 8,628 adults with hypertension was used. The study explored various sociodemographic, behavioral, and clinical factors.Multiple ML models were trained and their performance was rigorously evaluated using performance metrics [9].

Results demonstrated that key depression predictors were identified as income levels, triglyceride levels, white blood cell count, age, sleep disorders, and arthritis highly predictive ANN and SVM models [9].

The complex relationship between obstructive sleep apnea-hypopnea syndrome (OSAHS) and depression was investigated in [10] using NHANES data and machine learning (ML) such as logistic regression, the least absolute shrinkage and selection operator (LASSO) algorithm, and the random forest algorithm to predict depression risk. These were compared using metrics like receiver operating characteristic (ROC) area under the curve (AUC), specificity, sensitivity, and decision curve analysis (DCA) [10].

The logistic regression model outperformed the others, showing higher specificity and AUC values [10]. It pin-pointed key factors linked to higher depression risk in OSAHS patients, including gender, overall health, body mass index (BMI), smoking status, OSAHS severity, age, education, the ratio of family income to poverty (PIR), and asthma. A nomogram derived from the logistic regression model gives clinicians a visual tool to assess depression risk in OSAHS patients easily. This study proves the value of ML models in understanding and predicting depression among OSAHS patients [10]. By using large health datasets and advanced analysis, it sheds light on the intricate connection between sleep disorders and mental health.

III. Data Collection

The data was pooled from the 2005-2020 study cycle from the NHANES database published by the Centers for Disease Control and Prevention (CDC). NHANES is a population-based cross-sectional survey that aims to gather data on pertinent health, nutrition, diet, and behavior of American adults and children [1]. There were a total of twenty-eight questionnaire files containing information on participant demographics, lifestyle, depression, sleep disorder, and hypertension. The data in .xpt format was converted to .csv format for injection using pandas.

IV. Data Preprocessing

A. *Feature Extraction - Latent Features*

Feature 'Sleep_Apnea' was extracted from features 'How often do you snore' and 'How often do you snort/stop breathing?'. Instances where the latter had classes corresponding to values occasionally and frequently were considered to be affected by sleep apnea disorder, and were flagged as yes for the same.

Feature 'total_score' was extracted from the 9-item Patient Health Questionnaire-9 (PHQ-9) which utilized a four-point Likert scale. Each item was scored from 0 to 3, and the total score ranged from 0 to 27.

Feature 'depression_category' was extracted from features 'total_score' and 'Thought you would be better off dead'. Instances with a total score greater than or equal to 10 were categorized as 'severe depression', a total score greater than or equal to 5 as 'mild depression', and the rest as 'no depression'. In addition, instances wherein classes such as 'Several days', 'More than half the days', and 'Nearly every day' were present for the feature 'Thought you would be better off dead' were also categorized as 'severe depression.

B. *Data Cleaning*

All the column names were renamed to bring about consistency. Insignificant columns such as 'SEQN' and 'total_score' were dropped. Classes such as 'Refused' and 'Don't know' corresponding to various features were considered null values and were imputed using various methods listed below.

C. *Missing Value Imputation*

Two continuous features had missing values, namely 'PIR' and 'No_of_hours_you_sleep'. Median was used for imputing these. Figure 1 shows the distribution of these features before and after imputation.

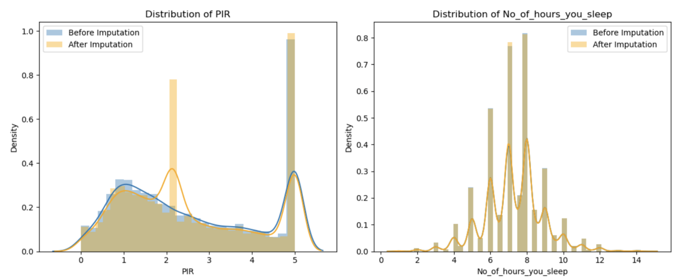

Fig. 1. Continuous features distribution before and after missing value imputation

The mode was used for imputing missing values for features that had discrete (16) and categorical (3) features. Figure 2 shows the count of each class for categorical features before and after imputation.

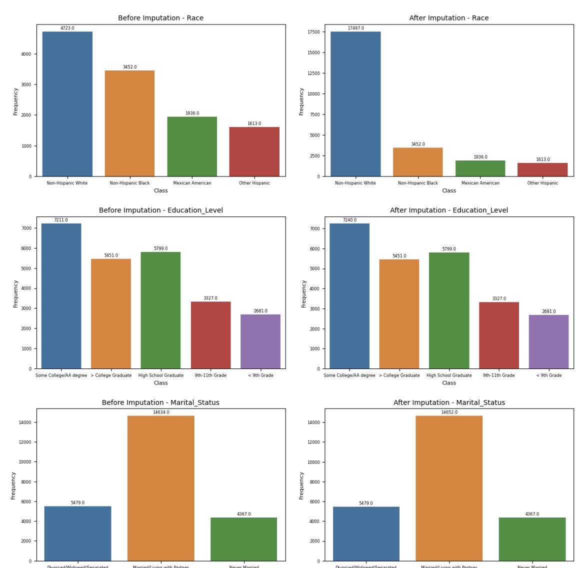

Fig. 2. Class counts for categorical features before and after missing value imputation

D. Outlier Detection and Treatment

Interquartile range (IQR) was used to detect and treat outliers in the data. Only one feature 'No_of_hours_you_sleep' had outliers. Value less than the lower limit and greater than the upper limit were considered outliers and removed. Figure 3 shows the distribution and boxplot before and after outlier treatment.

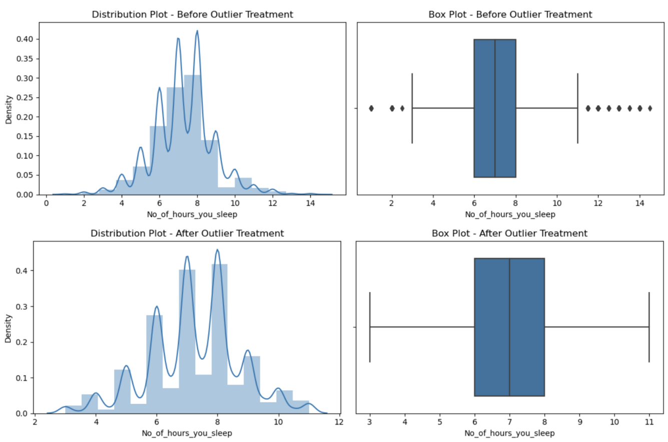

Fig. 3. Distribution and boxplot for 'No_of_hours_you_sleep' before and after outlier treatment

E. Feature Encoding

Feature encoding was done to convert categorical features to numerical ones. Label encoding was performed on ordinal features such as 'Education_Level' and 'depression_category'. One-hot encoding using get_dummies method was applied to nominal features such as 'Race' and 'Marital_Status'.

V. Data Analysis 1

The dataset comprises 24,108 observations, with a wide range of variables covering depression symptoms, sleep issues, hypertension, and demographic characteristics. Symptoms like "Little interest in doing things" and "Feeling down or depressed" have means of 0.34 and 0.32 respectively, suggesting that a significant number of respondents reported experiencing these symptoms, albeit at a lower intensity on average. Considering Sleep Issues, on average, respondents reported "Trouble sleeping or sleeping too much" with a mean score of 0.57, indicating that sleep-related issues are somewhat prevalent.

For Hypertension, a significant portion of the respondents reported taking medication for high blood pressure (HBP), with the most common response being '1', indicating that they are on medication. The dataset includes a diverse set of participants, with "Non-Hispanic White" being the most common race and "Married/Living with Partner" being the most common marital status. Gender is almost evenly split, and the majority of respondents have an education level of '3'.

Sleep Apnea and Hypertension: The majority of respondents reported not having sleep apnea (21,754 out of 24,108), and a majority are taking medication for high blood pressure (23,163 out of 24,108), which suggests a potential area of interest in exploring the relationship between hypertension and sleep apnea.

Lifestyle Factors: Most respondents are not current drinkers or smokers, which provides an interesting perspective on lifestyle factors that could be associated with depression, sleep apnea, and hypertension.

The correlation heatmap (Figure 4) provides a visual summary of the relationships between depression, sleep apnea, hypertension (blood pressure), and sleep duration. There is a modest positive correlation (0.02) between sleep apnea and the depression category, suggesting that individuals with sleep apnea might have slightly higher depression scores. The correlation between sleep apnea and blood pressure is very low (0.01), suggesting little to no direct relationship between these two variables in this dataset.

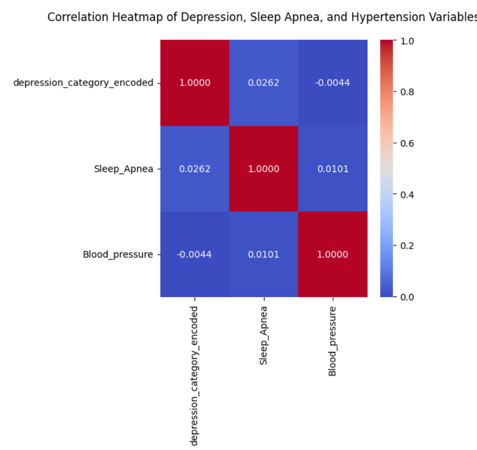

Fig. 4. Correlation Heatmap on Vital Features

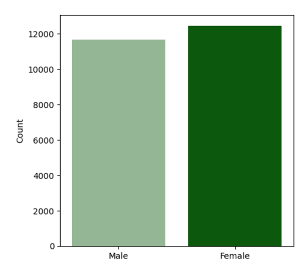

Fig. 5. Gender Distribution in Dataset

The gender distribution as seen in Figure 5 is a relatively balanced representation of male and female participants in the dataset. This balance is crucial for ensuring that any findings related to depression, sleep apnea, and hypertension can potentially be generalized across genders. A balanced gender distribution also allows for a more nuanced analysis of gender-specific trends or differences in the health conditions being studied.

Further exploration suggests a complex relationship between gender and depression. Females report higher rates of both mild/moderate and severe depression than males across the sample. This highlights the need for gender-specific mental health considerations and potentially tailored support systems.

The age distribution revealed a wide range of ages among participants, from young adults to the elderly. Figure 6 distribution is somewhat right-skewed, indicating a larger representation of middle-aged and older adults. This skewness is relevant because the prevalence of conditions like hypertension and sleep apnea can increase with age. Moreover, understanding age distribution is essential for analyzing depression scores, as age can influence both the growth and reporting of depressive symptoms.

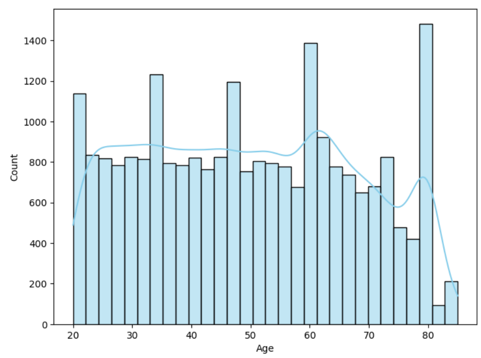

Fig. 6. Age Distribution in Dataset

Figure 7 illustrates how sleep duration varies across different depression categories. Some key observations include:

- Similar Median Sleep Hours: The median number of sleep hours appears relatively consistent across depression categories, suggesting that average sleep duration might not drastically differ by depression severity.
- Variability and Outliers: There's noticeable variability in sleep hours within each category, especially with outliers indicating significantly shorter or longer sleep duration. This variability might reflect the complex relationship between sleep and depression.
- Slight Differences in Distribution: While the medians are similar, the distribution and range of sleep hours show some differences. For example, the "Severe" depression category appears to have a slightly broader interquartile range, suggesting greater variability in sleep patterns among individuals with severe depression.

Considering Depression vs Sleep Apnea: If respondents are diagnosed with sleep apnea problems, it does not necessarily mean they will have depression as shown in Figure 8. The respondents without sleep apnea problems could still suffer from depression.

While studying Depression vs Hypertension, it was found that though there is a connection between high blood pressure and depression, they don't always occur together as reflected in Figure 9. Some people have high blood pressure without depression, and some people have depression without high blood pressure.

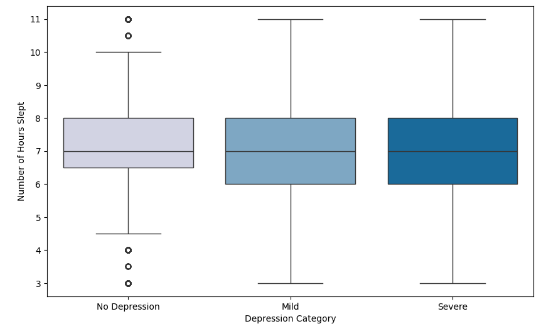

Fig. 7. Hours of Sleep by Depression Category

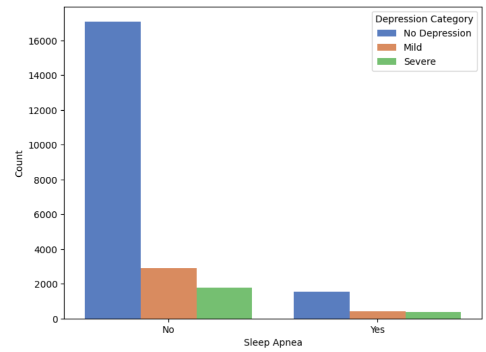

Fig. 8. Depression vs Sleep Apnea

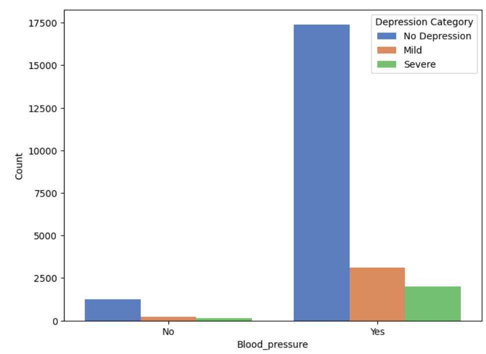

Fig. 9. Depression vs Blood Pressure (Hypertension)

The plot in Figure 10 illustrates the age distribution within each depression category. It shows that while the age distribution is somewhat similar across categories, there might be a slightly wider spread in the "No Depression" and "Mild" categories, indicating a broader age range. The "Severe" category appears to be more between the age range 20 to 66 ages suggesting specific age groups may be more prone to severe depression symptoms.

Interestingly, higher education levels correlate with the greatest percentage of individuals reporting no depression, suggesting a potential connection between education and mental well-being. However, it's important to remember that this data doesn't prove a causal link, and factors like healthcare access and socioeconomic status could also play a role.

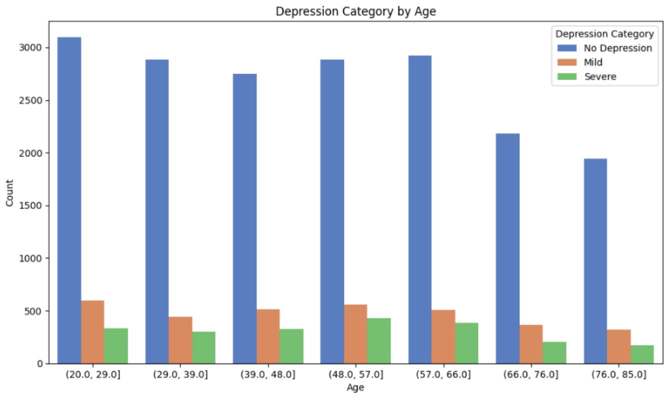

Fig. 10. Distribution of Depression Category by Age

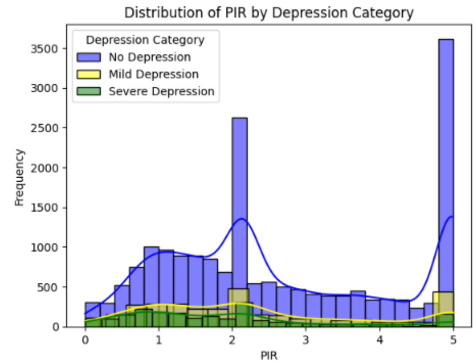

Fig. 11. Poverty to Income Ratio by Depression

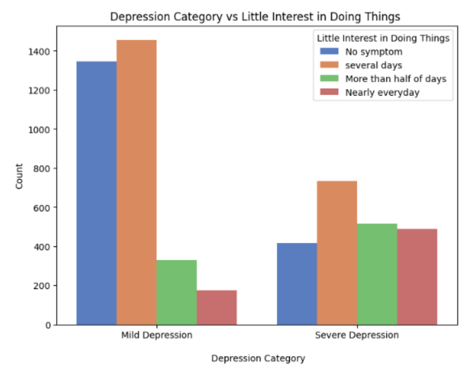

Fig. 12. Depression and Interest in Participation

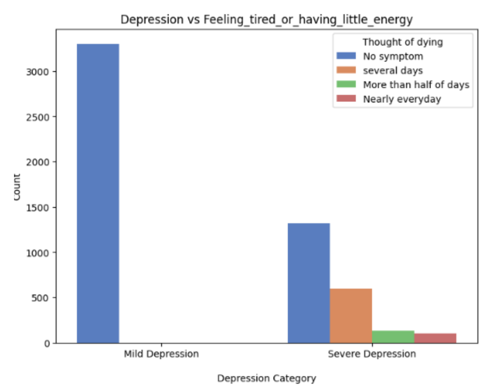

Fig. 13. Depression and Death thought

Regarding race, while non-Hispanic whites show the highest percentage of both no depression and severe depression, this finding needs further context. It may reflect population size, differing stressors, or disparities in healthcare access within the dataset.

Marital status also appears influential. Married or cohabitating individuals report the lowest rates of depression. Conversely, those who are divorced, widowed, or separated show higher rates of mild and severe depression. This emphasizes the potential impact of social support and relationship changes on mental health.

Overall, these findings offer valuable starting points for allocating resources and designing mental health programs. It's crucial to consider the complex ways education, race, and social circumstances interact with the experience of depression.

VI. Data Analysis 2

Although Data Analysis 1 suggests a foundational idea of what could be the expected factors responsible for depression in patients, it is not clear whether the same factors are responsible for depression. Hence, we have to dig deeper to find the specific factors that are affecting people with depression. For this, we explore different questions and possibilities that may be the cause of depression.

First, it was analyzed whether poverty to income ratio of family results in depression as shown in Figure 11.

As the data is highly imbalanced, patients with mild and severe depression have been considered to ascertain the actual factors that are causing the depression for further analysis.

The patients answered the 9 questions related to the depression based on four options: '0' indicating no symptom, '1' indicating they had a symptom several times over 2 weeks, '2' indicating symptoms more than a week over 2 weeks, and '3' indicating symptoms almost all days.

The distribution of depression by the No interest as seen in Figure 12 shows that many people showed no interest in doing things when diagnosed with depression over the past 2 weeks. A patient who had a thought of death have severe depression when compared to the patients with mild or no depression as shown in Figure 13.

Higher patients felt low or depressed almost all days in 2 weeks when they had severe depression compared to mildly depressed patients as shown in Figure 14.

Figure 15 shows the count of the number of symptoms each patient had experienced in the past 2 weeks when diagnosed with depression. It illustrates that most people experienced 3-5 symptoms in case of mild depression and 6-8 symptoms were experienced by patients with severe depression.

Figure 16 shows the number of patients with each symptom. More people experienced no interest in participation, feeling low, sleeping disturbances, and poor appetite symptoms over the past 2 weeks of the survey/examination.

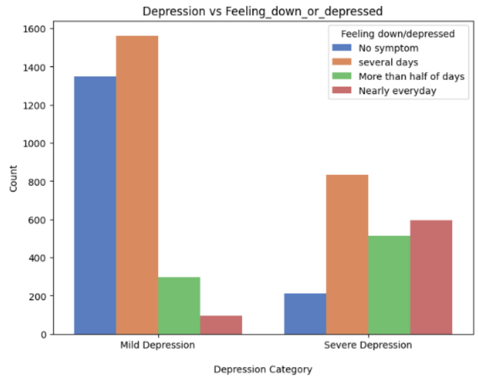

Fig. 14. Depression and Feeling low/depressed

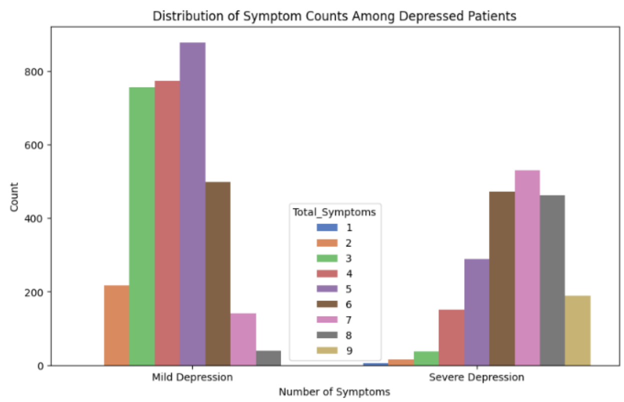

Fig. 15. Depression and Number of symptoms in a patient

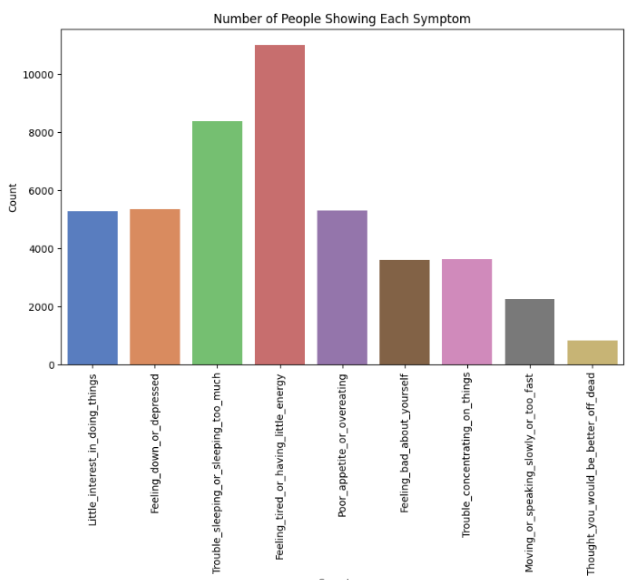

Fig. 16. Depression and Symptom

VII. Discussion and Community Impact

Through detailed examination, the current study has explored the complex interactions among depression, hypertension, and sleep apnea. Findings imply that different symptoms expressed by the patients can be used in analysing the depression in patient along with sleep apnea and hypertension for different age and gender. Notably, the usage of analytical techniques has identified a variety of parameters, including biological and social factors, indicating the complexity of the link between depression and hypertension. Likewise, the correlation between depression and sleep apnea is demonstrated, highlighting the reciprocal relationship between sleep disorders and mental health issues.

A gender-specific analysis indicated that females were more likely than males to experience depression, indicating the need for gender-sensitive mental health interventions. Age distribution analysis revealed a higher representation of middle-aged and older adults in the dataset, which is consistent with the notion that these age groups may be more susceptible to hypertension and sleep disturbances, which may affect the onset and course of depression symptoms.

The project's use of data mining techniques to predict depression offers a viable method for early identification and intervention that may ultimately help those with sleep apnea or hypertension. Healthcare professionals can more effectively treat the trifecta of depression, hypertension, and sleep apnea by identifying potential risk factors and implementing preventative screening procedures.

The study also emphasizes the necessity of a multifaceted approach to mental health that takes into account all of the interrelated elements that have an impact on an individual's well-being. The impact of lifestyle factors, such as smoking and alcohol use, was shown to be less significant, which led to a reassessment of their significance in connection to depression and related comorbidities.

VIII. Conclusion

This study emphasizes how important it is to understand and consider the relationships between long-term medical issues and mental health. Healthcare providers can identify high-risk patients and take prompt action by uncovering underlying risk factors using data analytics. The correlations shown between depression categories and the poverty income ratio indicate that socioeconomic status is a critical component of mental health and that initiatives targeted at reducing poverty may potentially have a positive impact on mental health outcomes.

The project culminates by highlighting the significance of early detection and intervention and offering a comprehensive perspective on depression in the setting of hypertension and sleep apnea. The study's technique, which makes use of data mining algorithms and large-scale datasets, establishes a backdrop for further mental health research. Future research will concentrate on improving predictive models' accuracy in a range of demographics and broadening their application to incorporate more pertinent health factors.

References

1 Li, E., Ai, F., and Liang, C., "A machine learning model to predict the risk of depression in us adults with obstructive sleep apnea hypopnea

syndrome: a cross-sectional study," *Frontiers in Public Health*, vol. 11, p. 1348803, 2024.

2 Gomes, S. R. B. S., von Schantz, M., and Leocadio-Miguel, M., "Predicting depressive symptoms in middle-aged and elderly adults using sleep data and clinical health markers: A machine learning approach," *Sleep Medicine*, vol. 102, pp. 123–131, 2023.

3 Handing, E. P., Strobl, C., Jiao, Y., Feliciano, L., and Aichele, S., "Predictors of depression among middle-aged and older men and women in europe: a machine learning approach," *The Lancet Regional Health–Europe*, vol. 18, 2022.

4 Lee, C. and Kim, H., "Machine learning-based predictive modeling of depression in hypertensive populations," *PLoS One*, vol. 17, no. 7, p. e0272330, 2022.

5 Sau, A. and Bhakta, I., "Predicting anxiety and depression in elderly patients using machine learning technology," *Healthcare Technology Letters*, vol. 4, no. 6, pp. 238–243, 2017.

6 Wei, T., Zeng, C., Chen, L., Wang, S., Li, S., Chen, Q., and Wang, L., "Predicting factors of depression in patients with primary hypertension: a community-based study," *International Journal of Psychiatry in Clinical Practice*, vol. 8, no. 4, pp. 213–217, 2004.

7 Van Ryswyk, E., Mukherjee, S., Chai-Coetzer, C. L., Vakulin, A., and McEvoy, R. D., "Sleep disorders, including sleep apnea and hypertension," *American journal of hypertension*, vol. 31, no. 8, pp. 857–864, 2018.

8 Hayley, A. C., Williams, L. J., Venugopal, K., Kennedy, G. A., Berk, M., and Pasco, J. A., "The relationships between insomnia, sleep apnoea and depression: findings from the american national health and nutrition examination survey, 2005–2008," *Australian & New Zealand Journal of Psychiatry*, vol. 49, no. 2, pp. 156–170, 2015.

9 Lee, C. and Kim, H., "Machine learning-based predictive modeling of depression in hypertensive populations," *PLoS One*, vol. 17, no. 7, p. e0272330, 2022.

10 Li, E., Ai, F., and Liang, C., "A machine learning model to predict the risk of depression in us adults with obstructive sleep apnea hypopnea syndrome: a cross-sectional study," *Frontiers in Public Health*, vol. 11, p. 1348803, 2024.