Digital Health and Human Behaviour - Project

Table of Contents

1. Introduction

Depression is a common mental disorder affecting 340 million people globally. The relationship between Depression and physical activity is well studied, with strong evidence suggesting that physical activity has a positive effect on depression symptoms comparable to that of conventional antidepressants and cognitive behaviour therapy treatments (Dinas et al, 2010). While this shows higher activity levels to be beneficial to counter depressive symptoms, less research has been done on to what extent the level of activity affects the likelihood of being diagnosed with depression.

Other factors also affect the likelihood of being diagnosed with depression. For instance, research has found that across different countries, “Women, unmarried individuals, and people with lower incomes are significantly more depressed” (Inaba et al., 2005). Since education is directly related to the income of an individual, it also influences a person’s mental health outlook.

It is also hypothesised (Smagula et al, 2021) that the timing of activity during the day can have an impact on depression symptoms. It was found that depression occurs at higher rates when the average activity of a person is low, and that being active in the mornings is especially important for preventing depressive symptoms.

Nowadays, wearable sensors are getting increasingly popular. This data can be used for mental health research. This data offers potential to gather information on individuals such as sleep patterns and exercise intensity. However, Since the data collected by wearables is non-public, it makes it difficult for researchers to collaborate and use this data for scientific research (Garcia-Ceja et al., 2018).

Understanding the effect of social factors and daily habits on mental health is critical for developing targeted interventions to prevent and fight mental disorders. By investigating patterns in this data set, I aim to uncover valuable insights that can help understand strategies to maintain good mental health.

The researcher question addressed in this report is: “How do daily activity levels relate to the presence and severity of depressive symptoms in a non-clinical population, and what are the potential implications for mental health interventions?"

In this exploration a multi-faceted approach was used, leveraging both traditional statistical analysis as well as machine learning techniques. The methods used are Statistical measures such as t-statistics and Cohen’s d, as well as Logistical regression, K-nearest neighbours, and Random Forest for classification. The group-level analysis revealed a statistically significant and practically meaningful difference in average activity levels between the control and condition groups. The logistic regression model exhibited the highest accuracy (82%) in classifying depression status, surpassing the K-Nearest Neighbors model (64%). The Random Forest model, classifying depression types, demonstrated an accuracy of 75%, with average activity identified as the most influential feature. It was found that activity levels throughout the day might have a significant impact on depression outlook of a person. It was also found that the kind of depression a person has can be determined through their average activity, their MADRS score, and their age.

2. Problem Formulation

The objective for this project is to study the effect of activity levels (sleep and exercise) on being diagnosed as depressed. The project further aims to investigate the relationship between age, marital status, employment, etc on the severity of depression. This is done by using the characteristics of the condition group, such as their MADRS scores and other social factors as well as their activity levels.

Thus, the main problem to be addressed in this report is to compare the activity levels of control and condition group to determine whether activity levels of an individual effect their likelihood of being depressed. Another goal is to study the condition group and evaluate which social factors result in which Afftype of depression. These social factors include factors which are studied already such as gender, age, marriage, and work as well as less researched factors such as self-reported melancholy and inpatient status. The affect of the activity levels during different times of the day on depression outlook will also be studied.

This is essential to gain insight into preventative strategies and understanding which socio-demographic factors put individuals at the most risk. By exploring these aspects, I aim to identify potential preventative strategies and shed light on overlooked factors that may significantly impact an individual's susceptibility to depression.

3. Dataset Description

For this project, the dataset consists of observations of patients with depression (23 patients) and a control group consisting of participants without depression (32 patients) as well as their activity levels. The activity levels are measured with an Actigraph watch which made measurements every minute lasting a period of 5–20 days. Each datapoint corresponds to a participant and includes information on the participant collected through the actigraph as well as other social features for the condition group.

The dataset was originally used to study motor activity of participants diagnosed with schizophrenia and other depressions (Garcia-Ceja et al., 2018). It consists of 55 patients as datapoints, each with three columns: timestamp (object), date (object), and activity. (int64). The “activity” is assessed using the readings made by accelerometer on the wearable, and its unit is “counts”. Each count corresponds to a voltage which is found by combining intensity, quantity, and duration of movement across all direction (Garcia-Ceja et al., 2018).

Furthermore, the dataset provides other information, which can be found in table 1. Column “afftype” refers to the diagnosis given to the patient. This can be Unipolar, Bipolar I, or Bipolar II. According to DSM-5, Unipolar disorder is characterized by persistent depressive symptoms while bipolar disorder involves mania (American Psychiatric Association, 2013). Specifically, Bipolar involves I intense maniac episodes, whereas in bipolar II disorder, the mania is less intense, also known as Hypomania.

The columns “Madrs1” and “Madrs2” show the MADRS score of patients at different stages of the study. MADRS (Montgomery-Asberg Depression Rating Scale) is a prevalent rating scale used to grade the severity of depression (0 to 60) based on observation and conversation with the patient (Garcia-Ceja et al., 2018).

Information about Afftype, Melanch, Inpatient, Edu, Marriage, Work, and MADRS scores is only present for the condition group.

Column	Description	Data Type
Number	Patient identifier	object
Days	Number of days of measurements	int64
Gender	Gender (1 for female, 2 for male)	int64
Age	Age groups	object
Afftype	Diagnosis (1: bipolar II, 2: unipolar depressive, 3: bipolar I)	float64
Melanch	Presence of melancholia (1: yes, 2: no)	float64
Inpatient	Patient status (1: stayed at a hospital, 2: outpatient)	float64
Edu	Education grouped in years	object
Marriage	Marital status (1: married or cohabiting, 2: single)	float64
Work	Employment status (1: working or studying, 2: not working)	float64
Madrs1	MADRS score when measurement started	float64
Madrs2	MADRS score when measurement stopped	float64

Table 1

3.1 Preprocessing:

While the dataset is comprehensive, there are a few datapoints which have missing values. This might be due to some information (like education and melanch) not being available. Since the ratio of such rows is small compared to the entire database, such rows were removed. Since the dataset is quite small, imputing missing cells was also considered. However, there is no way to impute the missing values, thus removing the rows was the better option.

The average activities of all the participants were calculated and a new column was added to the dataset to show the corresponding participant. Since the number of days each participant was monitored is different, we use the average to get a more general picture.

Furthermore, the hour from the datetime column for all control and condition participants was used to find the average activities throughout the day for each group.

The dataset has information about days, gender, and age for all participants. However, all other features are only for the condition group. Thus, only days, gender, age, and activity found from the corresponding actigraph dataset can be used to compare the control and condition groups.

Columns Gender, age, afftype, melanch, impatient, edu, marriage, and work were converted to category types. Since age and education were ranges of duration in years, for some purposes, it was easier to take the average of the corresponding age range and use that as the age for the participant. The cleaned dataset was split into control and condition datasets for better analysis. This was done since the number of features between the control and condition participants are different. Since there are two MADRS readings for the condition participants, another column called "average MADRS" was added to better compare the MADRS score between participants.

3.2 Preliminary Dataset analysis:

Figure 1 shows the distribution of all the participants in the dataset. As it can be seen, most participants have been monitored for 13 days, and both genders are represented well in the dataset. The ages of the participants have a big spread, with most participants falling under 45-49- and 50–54-year-old range. The participants are unevenly split in condition and control group, with 19 and 32 participants respectively. The distribution of the data is important to analyse, as it can help us understand the biases of the dataset and ensure that all conclusions consider the underlying distribution of the sample we are dealing with.

To further understand the distribution of condition group, a separate analysis on the control group was conducted. Figure 2 shows the distribution of participants in the condition group.

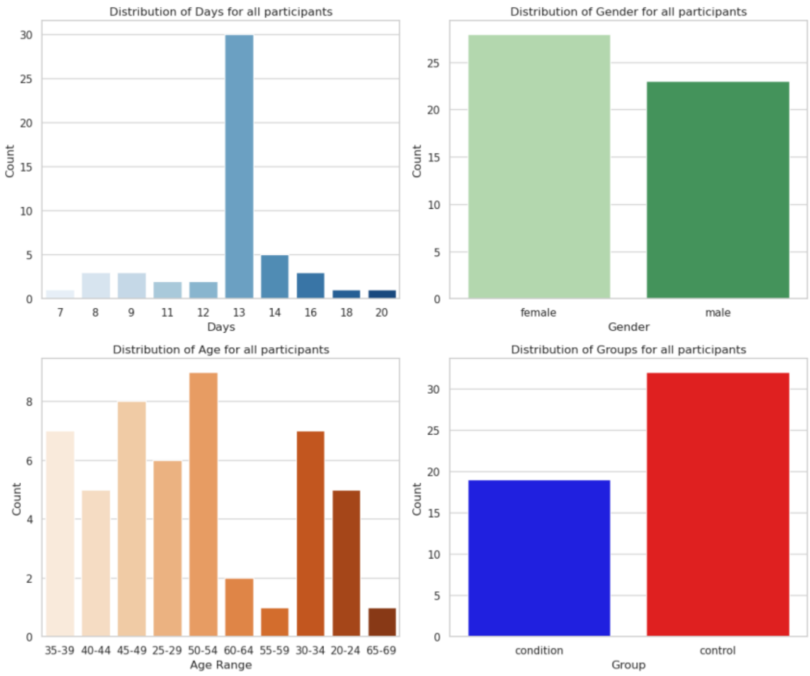

FIGURE 1 DISTRIBUTIONS FOR ALL PARTICIPANTS

As seen in Figure 2, the split between married and single participants as well as the males and females are almost identical. However, there is a big disparity between the amounts of participants with Melancholy and with none. Most participants are not working, and there is only one participant with Bipolar 2 depression. This uneven split makes it difficult to correctly identify patterns in the dataset, and different techniques will be used to reduce the effect of this uneven split.

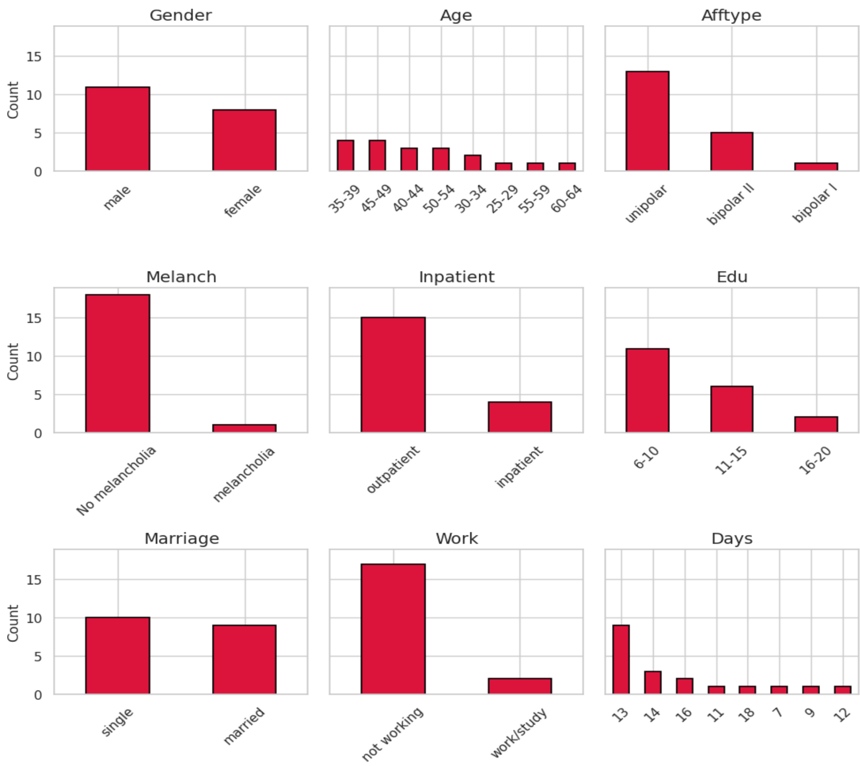

FIGURE 2 DISTRIBUTION OF THE CONDITION GROUP

To find any obvious patterns between MADRS scores and being a part of any demographic, the average MADRS score was plotted for each demographic (seen in Figure 3).

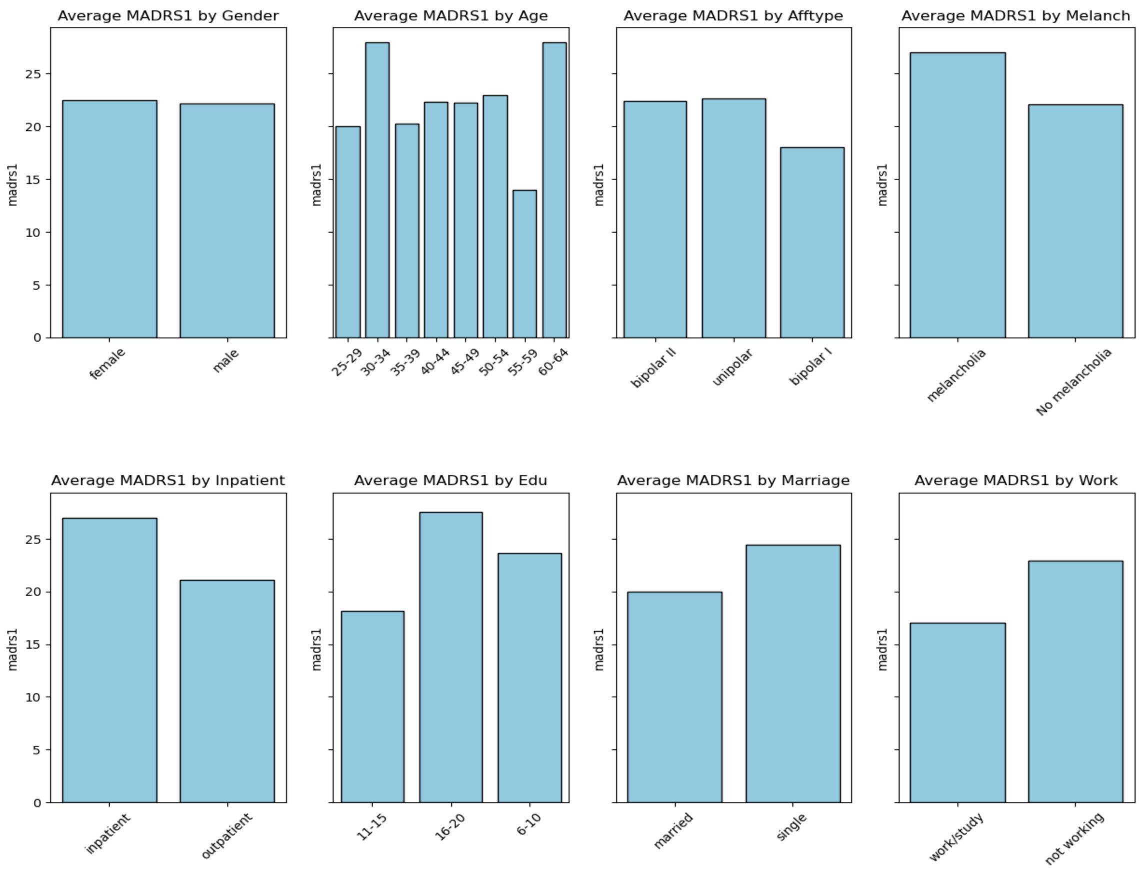

FIGURE 3 – MADRS SCORE BY DEMOGRAPHICS

3.3 Feature selection:

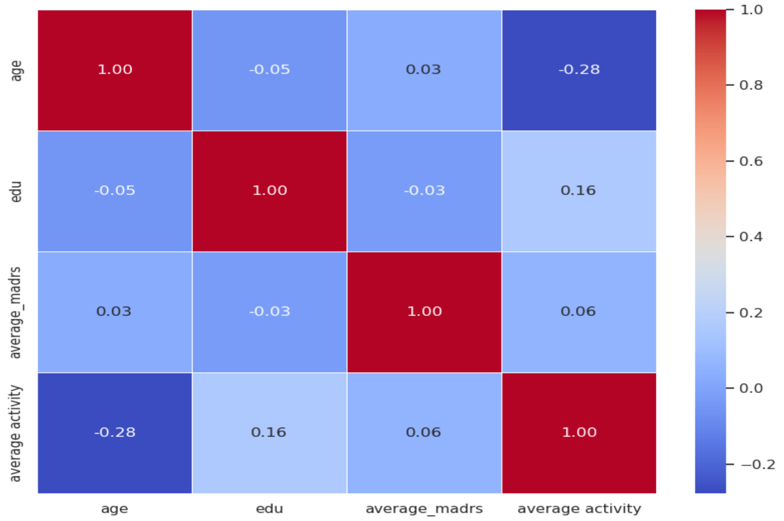

FIGURE A – CORRELATION BETWEEN FEATURES

As it can be seen in figure A, only "age" and "average activity" seem to have any kind of linear correlation. Thus, the correlation heatmap isn't very good at determining which features to choose. Hence, domain knowledge or other methods must be used for feature selection.

Since the report has multiple objectives, the features selected for each objective differ. For example, to compare the control and condition groups, the features selected are gender, age, and activity. Only these features are selected since the dataset does not have other features for the control group. The reasons to select gender, age, and activity as features is because according to domain knowledge, changes in these factors can affect the likelihood of being diagnosed as depressed. For instance, women are 1.7 times more likely to be diagnosed with depression (Albert, 2015). This could be explained by biological factors such as hormonal fluctuations in females during menstruation, pregnancy, and menopause, or through socioeconomic factors such as abuse, education, or income (Albert, 2015). However, the correlation between age and depression is not as simple. Some studies conclude that depression is less prevalent among old adults compared to their younger counterparts (Fiske & Wetherell, 2009), while others claim that the vast majority of depressed old people remain undiagnosed (Zenebe et al, 2021).

For comparing the severity and kind of depression for participants within the condition group, the features selected are Gender, Age, Melanch, Inpatient, Edu, Marriage, Work, and MADRS score.

The features selected for each model will be explained in greater detail in the methods section.

4. Methods:

4.1 Group Level Observations:

- ### Descriptive and Data analysis Methods:

One of the most straightforward ways to find a pattern between the activity levels and depression likelihood is to compare the average activity levels of the two groups. Figure 3 shows a violin plot comparing the average activity levels of the two groups.

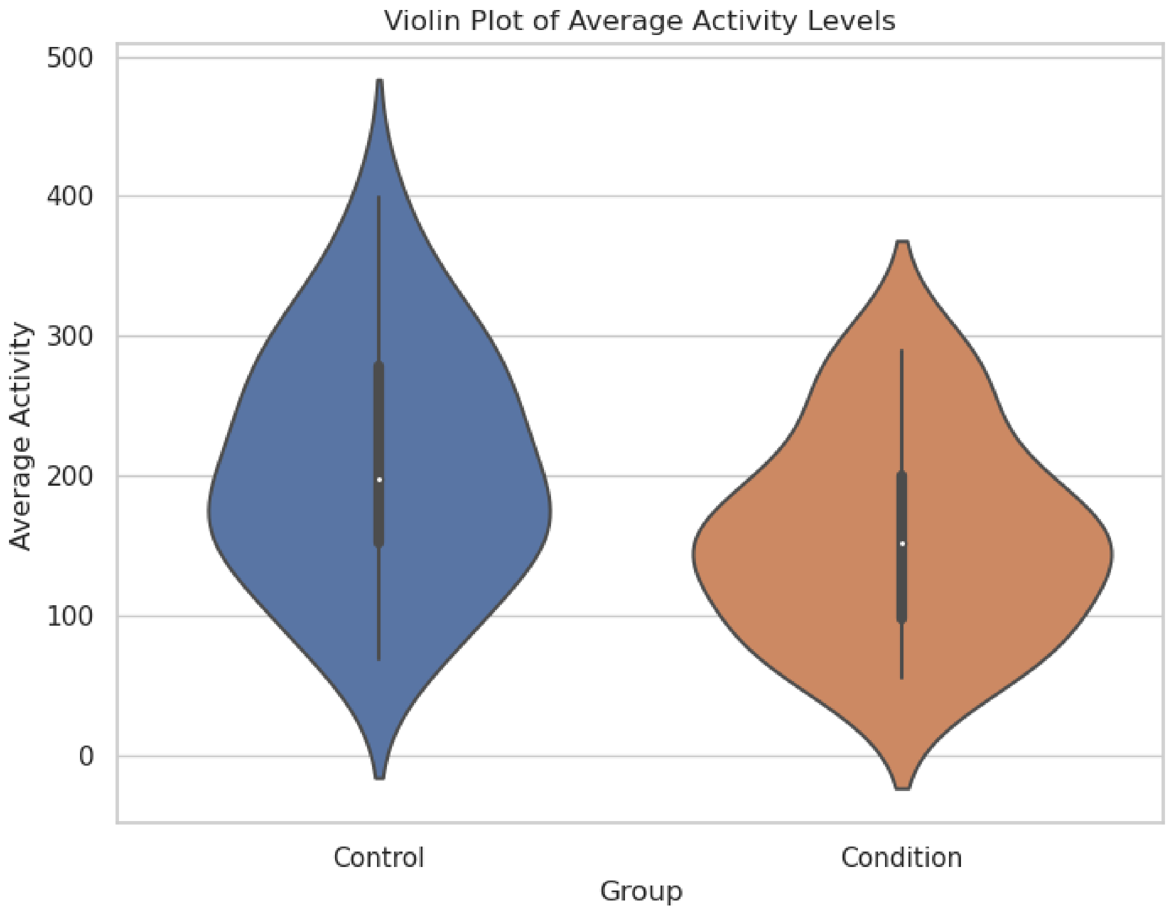

FIGURE 4 AVERAGE ACTIVITY LEVELS FOR CONTROL AND CONDITION GROUPS

Group	Mean	Std. Deviation	Minimum	25th Perc	Median	75th Perc	Maximum
Control	208.65	84.64	68.25	151.09	197.83	278.75	398.88
Condition	156.75	70.71	54.7	97.69	151.44	199.52	289.65

Table 2

As seen in figure 4 and Table 2, the control group has a higher mean activity level than the condition group, by over 50 points. The standard deviation is also higher for the control group, indicating that there is more variability in activity levels within that group. The minimum, 25th percentile, median, and 75th percentiles are all higher for the control group as well.

Thus, the violin plot and the data suggest that the control group is more active overall than the condition group. This increase in activity could explain why the condition group is diagnosed with depression, as according to many studies (Codella & Chirico, 2023), even small doses of physical activity can lower risks of Depression. However, this difference in average could also be due to the sample being biased. To check whether these results emerged out of sheer luck, other statistical tests are necessary. Thus, a T-test was conducted on the data to check whether the observed differences are statistically significant. In addition, the effect size (Cohen's d) was also calculated to measure how much the two groups differ when standardized. This is done by quantifying the magnitude of the difference between two groups in terms of standard deviations. This is advantageous as it helps compare the two groups with different sizes in a standardized unit (Lakens, 2013).

- ### Machine Learning Model 1 – Logistic regression to Classify Depression Status

A machine learning model was trained to classify the control and condition groups. The chosen method was Logistic regression. Logistic regression is a binary classification algorithm suitable for predicting the probability of an instance belonging to a particular class (Ranganathan, 2017). Logistic regression is justified in this situation since it is a well-suited algorithm for binary classification tasks where the goal is to classify instances into one of two classes. In this case, the objective is to classify participants into the "no depression" (control) or "depressed" (condition) groups.

The selected features for the model are 'gender', 'age', and 'average activity.' Categorical features ('gender' and 'age') were converted into numerical format using one-hot encoding to make them compatible with the logistic regression model. A new column "depression" was added to the database, which was found by checking which groups the participants were in (Control group was given "no depression" while condition group received "depressed"). This newly added column was the target variable (label). This contains the depression status of the participant.

The dataset consisted of 51 datapoints with 32 datapoints with "no depression" and 19 having the value "depressed". The dataset was split into training and testing sets, with 80% of the data used for training the model and 20% for evaluating its performance. Then the logistic regression model was created and trained using the training data. The trained model was used to make predictions on the test set. The accuracy of the model was calculated, and a classification report was generated. The classification report provides additional metrics such as precision, recall, and F1-score, offering a more detailed assessment of the model's performance in classifying participants into the "control" and "condition" groups.

- ### Machine Learning Model 2 – K-nearest neighbours to Classify Depression Status

Another machine learning model was trained for the same purpose. This was done to compare which model is more accurate in predicting the depression status of a participant. Herein, the model used was k-Nearest Neighbors (KNN) model. KNN is a classification algorithm that assigns labels to instances based on the majority class of their labelled k-nearest neighbors (Zhang, 2016). KNN was chosen as it can be invaluable for smaller datasets, and in this case, there are only 51 datapoints, so it was appropriate to use KNN.

In this case, the classifier was initialized with k=3, meaning it considers the labels of the three nearest neighbors to make predictions. Other values of K were considered, however k = 3, provided with the best accuracy to overfitting ratio.

The selected features and the labels are the same as the 1st model. The split used was also the same. To evaluate the model's performance, the accuracy was calculated using the accuracy score function. Additionally, a classification report was generated, providing detailed metrics such as precision, recall, and F1-score for each class.

- Machine Learning Model 3 – Random Forest to Classify Afftype among the depressed:

A machine learning model was trained to classify depressed people into Bipolar and Unipolar afftypes using a Random Forest Classifier. Random Forest is an algorithm which combines the output of multiple decision trees to reach a single result. According to Sharma (2023), Random Forest provides less overfitting and higher accuracy, is less sensitive to outliers and noise, and is more suitable for complex tasks with multidimensional input.

The features used were gender', 'age', 'melanch', 'inpatient', 'edu', 'marriage', 'work’, 'madrs1', 'madrs2', 'average_madrs', and 'average activity'. To use them appropriately in the random forest, age ranges were converted to average age and education levels were mapped to numerical values using a custom mapping (average). The label to be predicted was the combined afftype (Bipolar 1 and 2 were considered as one category, namely- Bipolar, and the other category was Unipolar).

The data was split into training and testing sets (80% training, 20% testing) to assess model performance. The preprocessing steps included standardizing numerical features and applying one-hot encoding to categorical features.

A pipeline, incorporating preprocessing and the Random Forest Classifier, was created, and the model was trained on the training set and evaluated on the testing set. A classification report was generated which provided a detailed performance assessment, including precision, recall, and F1-score for each mental health condition. Additionally, the “feature_importances” attribute of the Random Forest were used to assign how much effect each feature had on the classification.

- Comparing average activity by time of the day for control and condition groups:

As mentioned earlier, the time of activity can have a significant impact on depression outlook. To capture this effect (if it exists), visualising the average activity by the time of the day was done.

This helps to reveal patterns on trends in behaviour for both groups. A side-by-side comparison will help in understanding whether there are significant differences between the two groups.

Figure 5 shows the results of this comparison.

4.2 Subject Level Observations

- Analysing the most depressed person:

A subject level analysis on the most depressed person (Highest MADRS score) was conducted to find any interesting observations about that participant. The most depressed participant was found to be “condition_23”.

s	Gender	Age	Afftype	Melanch	Inpatient	Edu	Marriage	Work	Madrs1	Madrs2	Average Madrs	Average Activity
	F	30-34	Unipolar	No	Inpatient	16-20	Single	None	29.00	23.00	26.00	202.62

Table A – Characteristics of the most depressed participant

5. Results:

5.1 Results for descriptive statistics:

The following table shows the statistical measures corresponding to the average MADRS score across the two groups.

Statistical Measure	Value
T-Statistic	2.245
P-Value	0.029
Cohen's d Value	0.666

Table 3 – Statistics: average MADRS

As seen in table 3, t-statistic of 2.25 measures how far the sample mean of the condition group deviates from the sample mean of the control group, considering the variability in each group. A positive t-statistic indicates that the condition group's mean is higher than that of the control group.

The p-value of 0.029 represents the probability of observing the particular t-statistic as the one computed. This means that there is about a 3 percent chance that the data occurred by random chance. This is low enough to reject the null hypothesis – no difference between the two groups (Dahiru, 2008).

The Cohen's d value, approximately 0.67, suggests a moderate effect size (Lakens, 2013). A moderate effect size (0.67) implies that the observed difference is not only statistically significant but also practically meaningful.

These results indicates that the difference in average activity levels is not only statistically detectable but is also of a magnitude that may have relevance in real-world scenarios.

5.2 Results for Machine Learning Model 1: Logistic regression:

	Precision	Recall	F1-Score	Accuracy
Depressed	0.75	0.75	0.75	
No depression	0.86	0.86	0.86	
Total				0.82

Table 4 – Classification report For Model 1

Table 4 shows the classification report for Model 1. The overall accuracy achieved by this model is 0.82.

5.3 Results for Machine Learning Model 2: KNN:

	Precision	Recall	F1-Score	Accuracy
Depressed	0.50	0.75	0.60	
No depression	0.80	0.57	0.67	
Total				0.64

Table 5 – Classification report For Model 2

Table 5 shows the classification report for Model 2. The overall accuracy achieved by this model is 0.64.

5.3 Results for Machine Learning Model 3: Random Forest:

	Precision	Recall	F1-Score	Accuracy
Depressed	0.50	1.00	0.67	
No depression	1.0	0.67	0.80	
Total				0.75

Table 6 – Classification report for model 3

Table 6 shows the classification report for model 3. The overall accuracy achieved by this model is 0.75. The feature importance in the Random Forest can be seen in table 7. Other features had very low importance, and thus are not displayed here.

Feature	Importance
average activity	0.2205
madrs2	0.1894
average_madrs	0.171
madrs1	0.1289
age	0.0921

Table 7 - feature importance for random forest

5.4 Results for Comparison between control and condition activities throughout different times of the day:

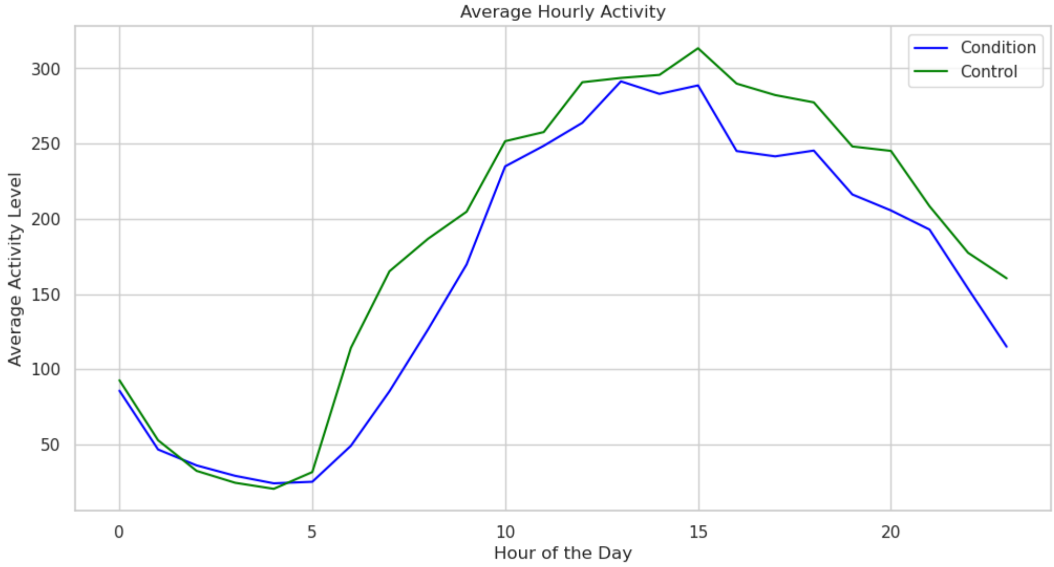

FIGURE 5 AVERAGE ACTIVITY LEVELS THROUGHOUT THE DAY

As it can been seen from the figure 5, the condition group remains lower active than the control group consistently after around 5 AM. The average hourly activity level for the condition group remains below that of the control group for the entire rest of the day, even though the difference between the two groups fluctuates over time. The difference in activity levels between the two groups is most pronounced in the morning hours. The activity level of the condition group gradually increases throughout until midday, but it never quite reaches the level of the control group. The activity level of the control group also fluctuates over time, but it generally remains higher than that of the condition group.

5.5 Results for Most depressed participant analysis:

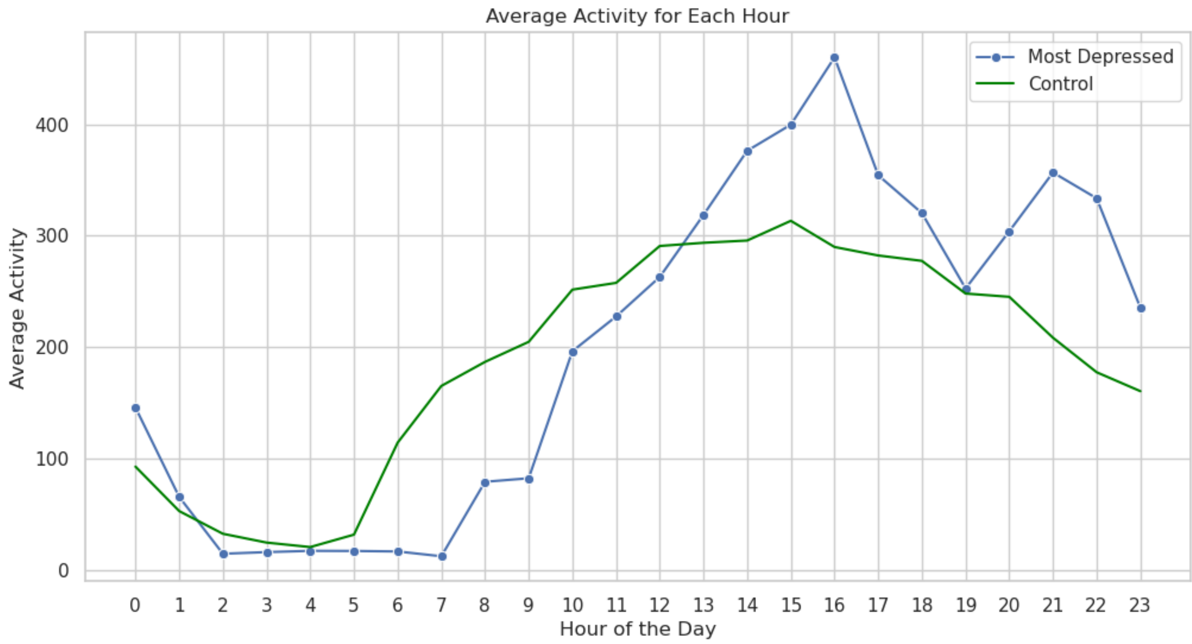

FIGURE 6 MOST DEPRESSED VS CONTROL AVERAGE

Figure 6 shows the average activity for each hour of the day for the most depressed subject (Condition 23) and the control group. As expected, the control group remains more active in the morning, which helps in fighting symptoms of depression. As it can be seen, both condition 23 and control group have similar activity levels from 0-5AM. However, after 5 AM the control group, on average, is much more active until around midday. These findings corroborate the hypothesis that being more active in the mornings tends to have a better prevention against Depression. It should be noted that although the peak activity level of depressed participant is much higher than the peak average control activity level, the average activity level for the individual was found to be 202.61, which is lower than the average control activity level (208.65). Thus, even if participant condition 23 has high levels of activity, which does provide some help against depressive symptoms, the distribution is what makes it more likely for them to have depressive symptoms.

- Results summarized:

The T-statistics and Cohen's d indicate a statistically significant and practically meaningful difference in average activity levels between the two groups.

Machine learning model	Accuracy
Model 1: Log Reg	0.82
Model 2: KNN	0.63
Model 3: Rand Forest	0.75

Table 8 – Accuracy for different ML Model

Since model 1 and model 2 both aim to Classify Depression Status, and model 1 has better accuracy, it can be said that model 1: Logistic regression, is better at predicting whether a person is depressed or not with an 82% accuracy.

The Random Forest demonstrated an overall accuracy of 0.75 in classifying whether a person with depression has Bipolar or Unipolar Afftype. In classifying participants, it was found that average activity was the most important feature in predicting the Afftype, followed by MADRS, and age.

Through time series analysis of the average activity, it was found that the condition group consistently exhibits lower activity levels, with the most pronounced difference observed in the morning hours.

6. Conclusions and Discussion:

In conclusion, the aim of the report is to compare the activity levels of control and condition group to determine whether activity levels and other social factors of an individual effect their likelihood of being depressed. It was found that the control group exhibited higher average activity levels compared to the condition group. Social factors such as gender, age, marriage, work, and education were considered in evaluating their influence on different types of depression (Unipolar, Bipolar I, Bipolar II). It was found that while using a random forest to classify these, the most important features are average activity, the MADRS score, and the age. Gender, Marriage, work, and education had seemingly very little impact or correlation with being depressed.

Time series analysis revealed that the condition group consistently exhibited lower activity levels, especially in the morning hours. The control group remained more active in the morning, potentially indicating a correlation between higher morning activity and a lower likelihood of depressive symptoms. Overall, the time-of-day analysis suggested that activity levels throughout the day might have a significant impact on depression outlook.

The report includes an analysis of the participant with the highest MADRS score (most depressed). It was found that despite having high activity levels, the participant had a very high MADRS score, suggesting that other factors might contribute to their depressive symptoms.

The results of the study have several implications, both in terms of understanding the relationship between physical activity and depression and informing potential interventions.

The study reinforces existing evidence that physical activity is linked to depression. The control group, with higher average activity levels, had a lower likelihood of being diagnosed with depression. The observed difference in average activity levels, especially in the morning, supports the idea that even small doses of physical activity, particularly at certain times, may have a protective effect against depression.

Knowing the type of depression can influence treatment strategies, and the identified features (average activity, MADRS scores, age) are crucial for predicting Afftype. The time-of-day analysis reveals consistently lower activity levels in the condition group, especially in the morning. Recognizing these temporal patterns can inform interventions that target specific times of the day for increased physical activity to mitigate depressive symptoms. Insights into the relationship between physical activity and depression, along with the influence of social factors, contribute to public health strategies. Developing targeted prevention strategies, considering individual characteristics and temporal patterns, can be more effective in reducing the prevalence and impact of depression. The challenges of using wearable sensor data for research, particularly data accessibility, underscore the need for collaboration and standardized data-sharing practices. Further research leveraging wearable technology can provide deeper insights into the dynamic interplay between physical activity and mental health.

Limitations and potential improvements: A comprehensive evaluation of the main limitations has been undertaken to enhance the transparency and reliability of our analysis. While the study provides valuable insights, I acknowledge the inherent limitations, such as the small dataset size and the fluctuations in number of recorded days for different participants. These limitations warrant caution in generalizing the findings to broader populations, and future research should address these constraints. A larger and more diverse sample would enhance the generalizability of the findings. The missing social data for the condition group also made it harder to compare the condition and control groups based on the social factors.

The uneven distribution between the control (32 participants) and condition (23 participants) groups might introduce bias. Furthermore, while wearable sensors offer valuable data, the dataset's reliance on Actigraph watches has limitations. Different types of wearables or additional sources of data (Heart monitor, GPS) could provide a more nuanced view of participants' behaviors. The analysis suggests that morning activity might be crucial, but it lacks a detailed investigation into circadian rhythms.

<u>Suggestions for Future analysis:</u> Conducting a longitudinal study with more MADRS tests being conducted could capture changes in participant's activity levels and mental health over time. Collecting more social information about both groups could also help in comparing the two groups. Moreover, integrating psychological factors such as stress, coping mechanisms, and personality traits into the analysis would contribute to a more holistic view of depression determinants.

Incorporate physiological data, such as heart rate variability, sleep patterns, and cortisol levels, to gain a more comprehensive understanding of the physiological aspects of depression.

To address the imbalance in control and condition participants to avoid skewed results, perhaps techniques like oversampling or advanced machine learning algorithms such as SMOTE that handle imbalanced datasets could be used.

7. References:

1. Albert P. R. (2015). Why is depression more prevalent in women?. *Journal of psychiatry & neuroscience : JPN*, *40*(4), 219–221. https://doi.org/10.1503/jpn.150205

2. American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™* (5th ed.). American Psychiatric Publishing, Inc[1]

3. Codella, R., & Chirico, A. (2023). Physical Inactivity and Depression: The Gloomy Dual with Rising Costs in a Large-Scale Emergency. *International journal of environmental research and public health*, *20*(2), 1603. https://doi.org/10.3390/ijerph20021603

4. Dahiru T. (2008). P - value, a true test of statistical significance? A cautionary note. *Annals of Ibadan postgraduate medicine*, *6*(1), 21–26. https://doi.org/10.4314/aipm.v6i1.64038

5. Dinas, Petros & Koutedakis, Yiannis & Flouris, Andreas. (2010). Effects of exercise and physical activity on depression. Irish journal of medical science. 180. 319-25. 10.1007/s11845-010-0633-9.

6. Fiske, A., Wetherell, J. L., & Gatz, M. (2009). Depression in older adults. *Annual review of clinical psychology*, *5*, 363–389. https://doi.org/10.1146/annurev.clinpsy.032408.153621

7. Garcia Ceja, Enrique & Riegler, Michael & Jakobsen, Petter & Torresen, Jim & Nordgreen, T. & Oedegaard, Ketil & Fasmer, Ole Bernt. (2018). Depresjon: A Motor Activity Database of Depression Episodes in Unipolar and Bipolar Patients. 10.1145/3204949.3208125.

8. Inaba, A., Thoits, P. A., Ueno, K., Gove, W. R., Evenson, R. J., & Sloan, M. (2005). *Depression in the United States and Japan: Gender, marital status, and SES patterns*. *Social Science & Medicine*, 61(11), 2280–2292[12]

9. Lakens D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

10. Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Logistic regression. *Perspectives in clinical research*, *8*(3), 148–151. https://doi.org/10.4103/picr.PICR_87_17

11. Sharma, A. (2023) *Random Forest vs decision tree: Which is right for you?, Analytics Vidhya.* Available at: https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/ (Accessed: 11 October 2023).

12. Smagula, S. F., Capps, C. S., & Krafty, R. T. (2021). Evaluating the timing of differences in activity related to depression symptoms across adulthood in the United States. *Journal of affective disorders*, *284*, 64–68. https://doi.org/10.1016/j.jad.2021.01.069

13. Zenebe, Y., Akele, B., W/Selassie, M. *et al.* Prevalence and determinants of depression among old age: a systematic review and meta-analysis. *Ann Gen Psychiatry* **20**, 55 (2021). https://doi.org/10.1186/s12991-021-00375-x

14. Zhang Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine*, *4*(11), 218. https://doi.org/10.21037/atm.2016.03.37