Aalto University
School of Science

An Analysis On Depression Using Actigraphy Data

1 Introduction

1.1 Context Setup

Depression, a severe mental disorder characterized by symptoms like sadness and anxiety, varies in severity based on the duration and impact of these symptoms on daily life. It frequently coexists with bipolar disorder, where manic episodes are a defining element, and both conditions have genetic components influenced by environmental factors. Contributing factors to depressive symptoms include disrupted biological rhythms, physical health issues, life events, and substance abuse (Garcia-Ceja et al., 2018).

In addition, modern wearable sensors are widely used to measure various aspects of individuals' activity. There is a growing recognition of the relevance of this activity data to various mental health concerns, such as fluctuations in mood and difficulties coping with daily challenges or stress within the field of psychiatry (Garcia-Ceja et al., 2018). Actigraphy is a method of monitoring motor activity that can be used to assess depression severity. Depressed patients often present with motor activity abnormalities that can be easily recorded using actigraphy (Peis et al., 2020).

This project aims to instill insight about depression by co-exploring actigraphs and other attributes among members of the controls and patients of the condition group.

1.2 Related Work

Over the past decade, there have been several studies around the use of actigraphy in depression research.
To begin with, (Razavi et al., 2011) discusses in their article how quantitative motor activity was assessed with 24-hour actigraphy recordings and argued that items such as retardation, agitation, and activities would correlate with actigraphic data. Similarly, (Peis et al., 2020) talks about how depressed patients present with motor activity abnormalities and how actigraphically recorded motor activity may predict inpatient clinical course and hospital discharge.

Moreover, Berle *et al.* in (Berle, Hauge, Oedegaard, Holsten, & Fasmer, 2010) discuss how motor activity patterns in patients with major depression are different from those of patients with schizophrenia and healthy controls, although motor activity was significantly reduced in both schizophrenic and depressed patients when recorded using wrist-worn actigraphs for two weeks.

1.3 Analysis Overview

Motivated by relevant efforts in this field, in this study we explore actigraph recordings of two groups of healthy and condition subjects in order to instil valuable insight about depression, differences in activity patterns, and finally propose two Machine Learning based approaches (classification and clusterings) that can help us detect this phenomenon using sensor data.
Our Decision Tree classifier reaches the accuracy of 73%. We also offer visual evidence that the sensor recordings can capture useful behaviors to some degree. Moreover, we extract various features from sensor data and fit a clustering model on them.

In the following sections, the formulation of the problem will be discussed, followed by a thorough assessment and introduction of the dataset. Next, our intended methods and motivations will be explained, and some numerical and graphical insights will be provided. Lastly, we conclude the article by elaborating on the results and their implications, the limitations, and the future directions imagined for the subject.

2 Problem Formulation

In this section we aim to set goals in a focused manner, emphasizing on previously less explored approaches, and leave the extra analyses to future researchers. There are a number of opportunities that this study aims to address:

(i) Making comparisons between control and condition groups and between different types of depression and providing meaningful plots from further observations. According to Figure 1, these comparisons are worth exploring.

(ii) Developing a system capable of classifying depression states and predicting MADRS score based on activity data.

(iii) Comparing sleep patterns of depressed vs non-depressed subjects. Also, more seasonal trends are expected to be observed during this analysis.

(iv) Evaluating different machine learning methods, such as classification techniques and clustering techniques to deal with dataset imbalance.

Figure 1 shows the relationship between depression level and personal characteristics. The sample count for each is not big enough to call the patterns statistically significant, but some reasonable trends are observed. That's the reason for including this part in our analyses.

3 Dataset

3.1 Data Collection

The data used in this project is obtained from the Depresjon dataset (Garcia-Ceja et al., 2018), which comprises motor activity recordings of the control and condition groups to study depression episodes in unipolar and bipolar patients.
Movement was measured using an actigraph watch positioned on the wrist, which quantifies activity levels by recording movements which exceed $0.05g$ at a sampling frequency

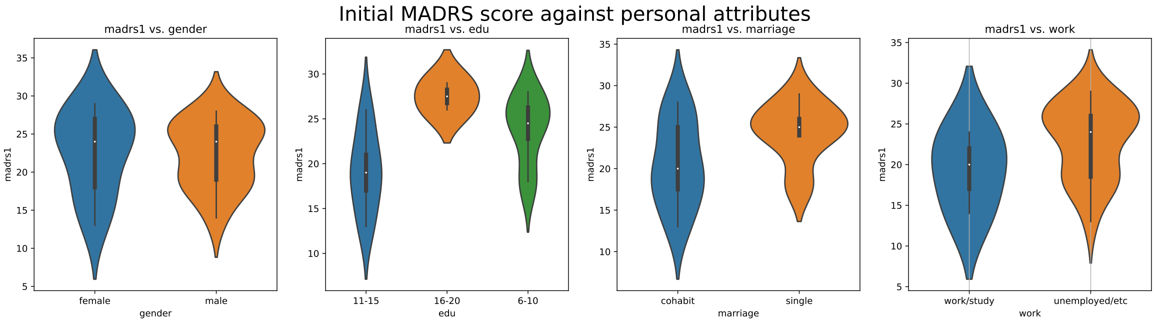

Figure 1: The violin plots show the distribution of MADRS score in the beginning of the study against personal attributes.

of $32Hz$. These movements generate a voltage, stored as an activity count in the watch, which is indicative of movement intensity. Total activity counts are consistently recorded at one-minute intervals (Garcia-Ceja et al., 2018).

3.2 Dataset Details

There are two groups containing high-resolution time series data of the motor activity recordings for the control and condition patients. The condition group consists of 23 unipolar and bipolar depressed patients, and there are 32 healthy subjects in the controls. The control group subjects' data is available for more than three weeks on average, whereas the condition group patients have mostly reported their data for around 17 days. However, the `days` column reports an average of 12.5 days for both groups for reasons unknown to us.

The *Score* data contains other attributes about the study subjects, including number of days of measurements, gender, age, affectivity type (bipolar II, unipolar depressive, bipolar I), melancholia, inpatient/outpatient, education (grouped in years), marital status, occupation status, and Montgomery-Åsberg Depression Rating Scale (MADRS) score when measurement started and when stopped). We also observe that three patients (more than 10%) among the condition group have NaN values in the `melanch` column (and 1 NaN value in `edu`). Two approaches could be either dropping the rows or the column in this regard.

3.3 Preprocessing and Feature Selection

Considering the class imbalance in the `melanch` and `inpatient` columns, the more reasonable approach would be dropping these columns. Other attributes are retained to highlight differences in further visualizations.

The resolution level in actigraphs is too much (i.e. minutes) and, therefore, not helpful in finding reliable patterns. It requires accumulation in larger bins (e.g. hours), which enables us to extract relevant features like chronotype. Also differences between weekdays and weekends require further attention. In addition, there are 10 age groups which can be narrowed down for better generalization. The same approach applies to MADRS levels which include five levels, mainly *1. normal*, *2. mild depression*, *3.moderate depression*, *4. severe depression*, and *5. very severe depression*. Figure 2 compares this score for patients in the beginning and the end of data collection, and also tries to explore the effect of age on their trends.

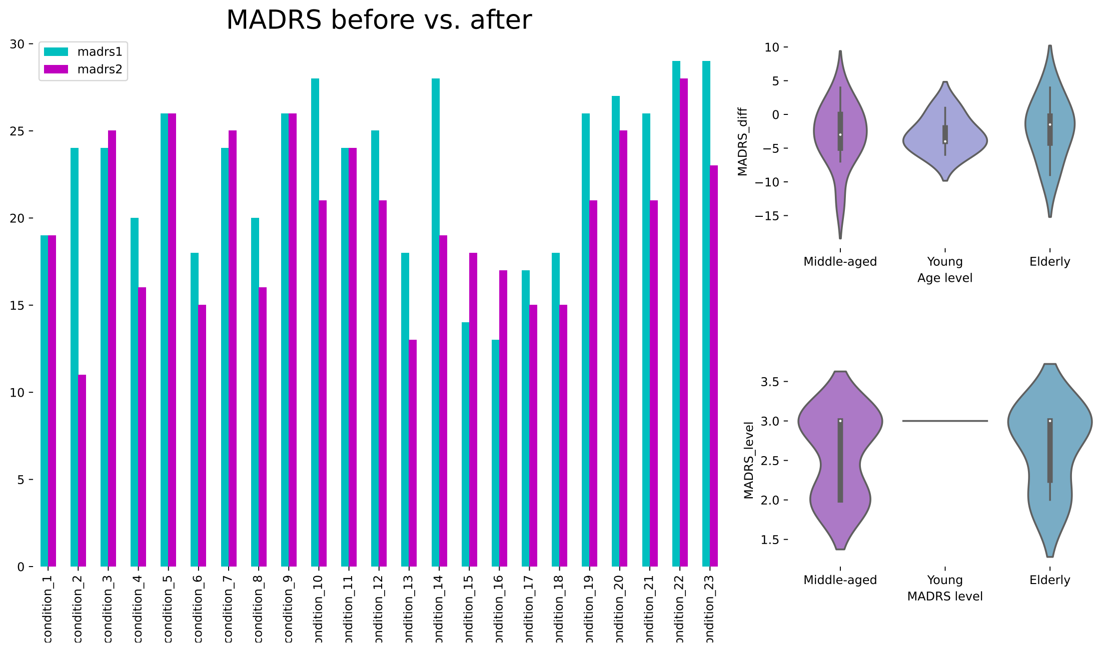

Figure 2: The left figure shows the difference in the MADRS level in the beginning and the end of the study, while the right violin plots explore the relationship between distributions of this change and levels with age groups

A significant step in the initial data exploration comprises extracting groups and meaningful aggregations from various pseronal attributes and activity data. This enables us to identify characteristics and behaviors that help make the data more interpretable, and also more separable for classifiers. Some examples of this approach include aggregating activity in different times of the day in different days of the week. Some general piece of information such as individual's average activity patterns which rely on our background knowledge (regarding activity levels among depressed people compared to healthy subjects) are also included in the initial feature list for further exploration.

4 Methods

In this section, we start by extracting useful information from actigraphs and make comparisons where relevant. Firstly, the samples are compared against each other and against overall trends. In the second part, we try to fit appropriate models on our data to recognize patterns both in supervised and unsupervised manners.

4.1 Individual Analysis

As mentioned in Section 3.3, the one-minute resolution in the activity recordings is not useful for our application, so we accumulate them in one-hour buckets. Figure 3 shows the average activity observed throughout the study.

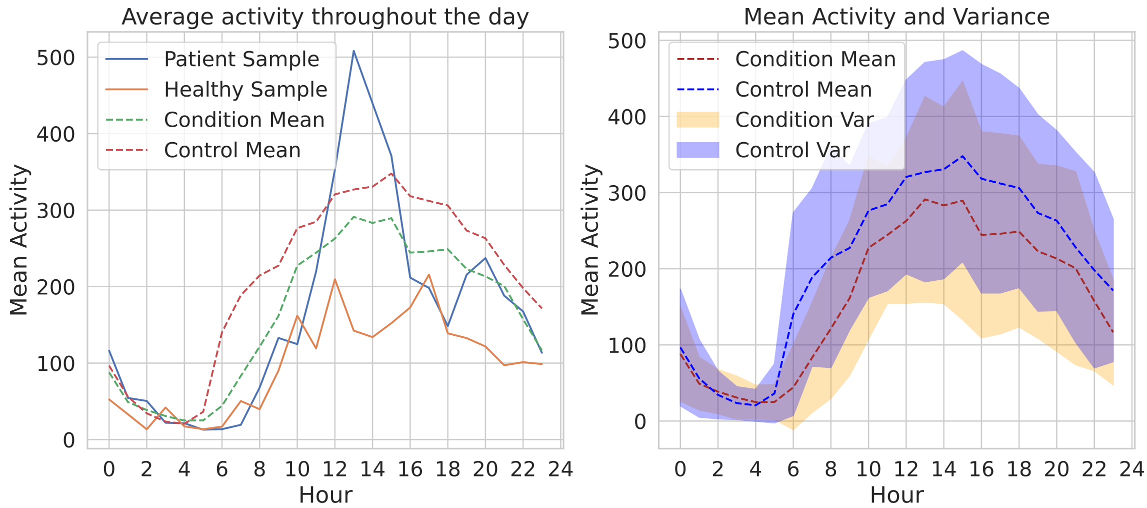

Figure 3: In the left plot, each line shows the average activity for a group or a random sample during 24 hours a day. In the second plot, variance is also drawn around the average lines.

4.2 Models

One aim of this study is to develop a model which is able to detect depression status using sensor recordings. To this end, we try both supervised methods i.e classification models, such as Decision Tree and Random Forest, and also K-Means clustering as an unsupervised method that tries to separate health subjects from patients.

4.2.1 Classification

In this part, we use classification models for two different tasks. Firstly, we train a classifier on categorical attributes of the condition group and using affective type as the label. And secondly, we train a classifier on both groups based on the features extracted from their activity data.

Affective type classification The classifier we use in this task is a Decision Tree. This is a reasonable choice in the context of a small dataset with few data points and features, because we have categorical features, and more importantly, we try to achieve interpretability and avoid overfitting that is highly likely while training on 22 rows.

Depression detection In this task, the main concern is extracting useful insights from the actigraphs to feed to the classifier. In this task we use a Logistic Regression model, because it's generally recommended to use simpler models that are less prone to overfitting when dealing with small datasets such as ours.
The extracted features from the activity data are the number of inactive hours on weekends and on weekdays, and average activity on weekends and on weekdays (four features in total for 55 rows).

4.2.2 Clustering and Dimensionality Reduction

In this section we select the suitable features and fit a K-Means model with two kernels to try to capture and distinguish the underlying distibution of each group.

In order to visualize the findings, a method for dimensionality reduction is required to map the high-dimensional data points to a lower-dimensional space i.e. the Cartesian plane. To do so, we utilize the t-distributed Stochastic Neighbor Embedding (t-SNE) method.

The results of these methods will be further discussed in the next section.

5 Results

In this section, both observations and models and their corresponding results are discussed.

5.1 Analysis Deductions

Group level observations Some patterns have already been introduced in previous sections. To elaborate more on the observation, Figure 1 suggests that individual characteristics can affect depression in a way. Also, Figure2 (top-right plot) shows that older groups are more likely to experience major sudden changes in their depression levels than younger people.

In addition, Figure 3 plots indicate that average activity level among the patients has always been lower than the healthy subjects, which sounds reasonable based on the nature of depression. Moreover, the variance area suggests that the deviation level among the control group is generally larger than depressed patients, especially in the early morning. However, no definitive markers are observed for different hours of the day. Therefore, studying chronotypes, which intuitively can be correlated to depression to some degree, cannot be performed with the existing data.

One more interesting goal is to identify different types of disorders based on this data, which is shown in Figure 4. This plot - although lacking enough data and accuracy - suggests that the *bipolar II* type aligns with moderate levels of depression.

Subject level observations One more observation concerns the patient with the highest MADRS score. By examining the average activity (shown in Figure 5), different anomalies are observed. The patient is twice as active compared to the condition group in the morning, but the activity level suddenly falls to almost no activity in the evening.

5.2 Classification Results

Affective type classification As mentioned in the previous section, we train a Decision Tree on our initial features to see if they show an underlying connection between the affective type and general attributes. Figure 6 shows the tree plot of this classifier that

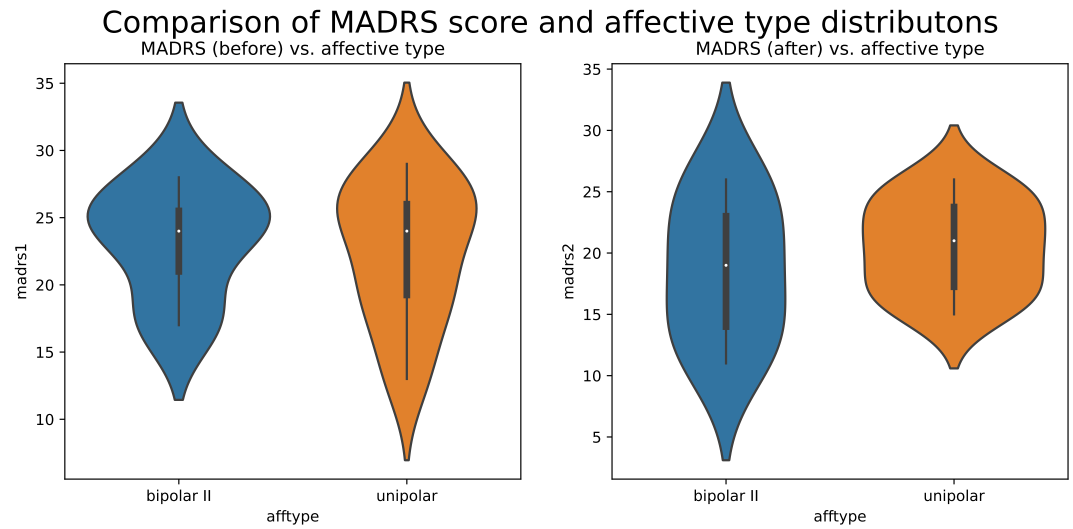

Figure 4: These violin plots compare the distribution of MADRS scores against affective types.

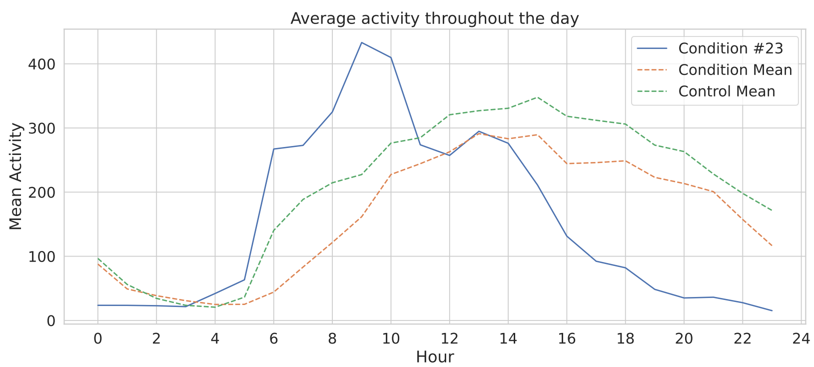

Figure 5: The mean activity level of an extreme case compared to the group average.

achieved a 60% accuracy, which is not reliable. However, apart from MADRS scores, marital status and age seem to be important factors.

Depression detection In this task, the main concern is extracting useful insights from the actigraphs to feed to the classifier. The Logistic Regression model trained on four activity-related features obtained a 73% accuracy, which looks promising and will be further explored in the next section.

5.3 Clustering

A 3D plot of the K-Means clustering on the same features as the previous classification task is shown in Figure 7.
The results for the K-Means clustering analysis is shown in Figure 8. In this figure, t-SNE has been used as a dimensionality reduction method to visualize the data.

The plot indicates that these features are not accurate enough to capture the underlying distibution of individuals' activity.

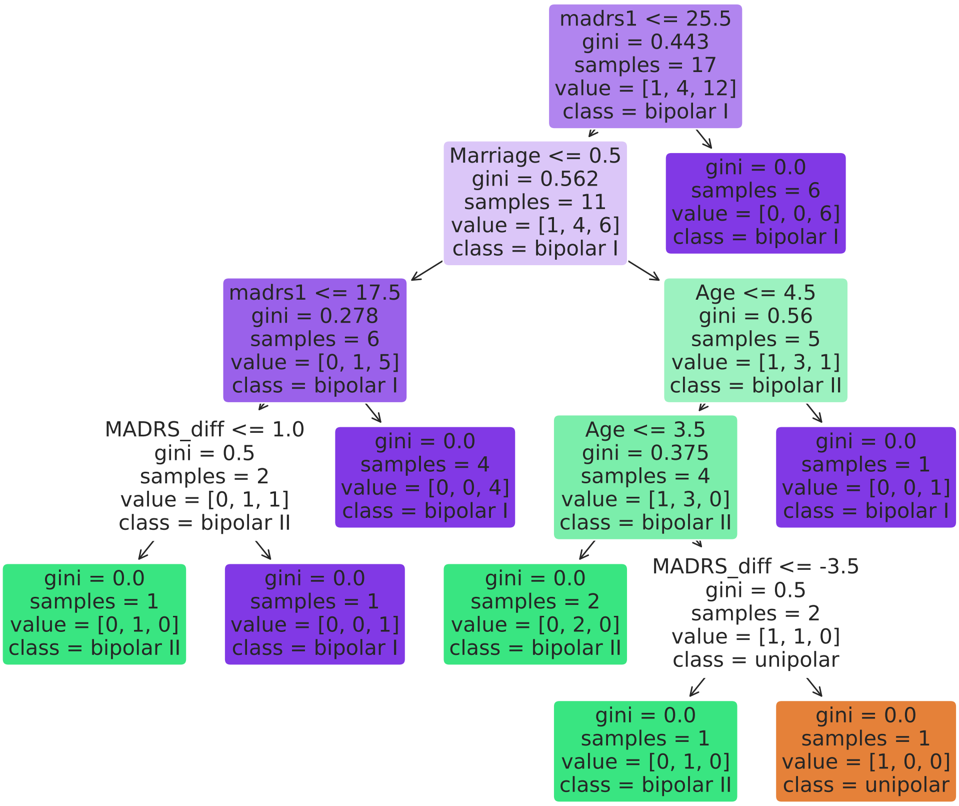

Figure 6: The tree plot indicates the decision process and feature importance in our model.

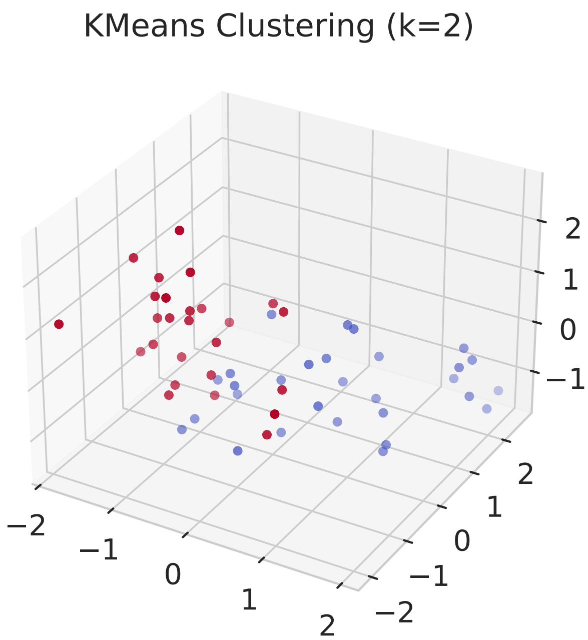

Figure 7: The 3D plot indicates the way our data points are scattered in space.

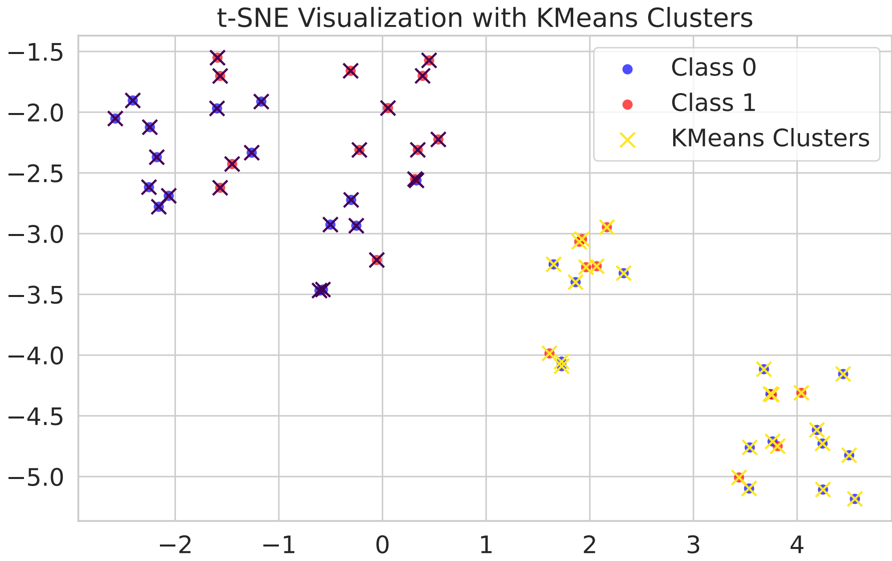

Figure 8: The t-SNE scatterplot with both real labels and K-Means labels.

The reason we choose t-SNE over other methods like Principal Component Analysis (PCA) is the ability to capture complex distributions that t-SNE offers through a probabilistic approach.

6 Conclusion & Discussion

The use of actigraphy and sensor data in the analysis of depression levels has provided valuable insights into the relationship between physical activity patterns and depression levels. In this study, we aimed at identifying patterns in patients by exploring their phenotypes and sensor data using visualization and machine learning methods. Our results suggest that not only there exists a certain relationship between depression and personal states e.g. work, education, and marital status, but also sleeping patterns are worth deeper analysis using advanced learning methods.

It is necessary to mention that our study was negatively affected by some limitations. The most important problem concerns the data quality and collection process. Apart from lack of accuracy and reliability issues that actigraphies have, the small number of patients and inconsistency in reporting the data adversely affects our classification accuracy. Such limitations become more vivid when interpreting actigraphy patterns to study depression.

The implications of this study for future research on depression and wearable devices are not significant, but we hope to have paved the way for future research in some key areas. For instance, the extracted features that were significantly different for the condition group, such as sleep patterns can serve as bases for more focused data collection. Researchers can also explore the effect of treating the root causes of such behaviors in pharmaceutical studies. Moreover, the classification models presented in this study can be valuable to future pathophysiological studies. Subsequent research can further explore and optimize machine learning algorithms for more accurate detection of depression using

sensor data.
Overall, the integration of wearable devices and sensor data in depression studies has a great potential for contributing to understanding this complex condition.

References

Berle, J. O., Hauge, E. R., Oedegaard, K. J., Holsten, F., & Fasmer, O. B. (2010). Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. *BMC research notes*, *3*(1), 1–7.

Garcia-Ceja, E., Riegler, M., Jakobsen, P., rresen, J. T., Nordgreen, T., Oedegaard, K. J., & Fasmer, O. B. (2018). Depresjon: A motor activity database of depression episodes in unipolar and bipolar patients. In *Proceedings of the 9th acm on multimedia systems conference.* New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/3204949.3208125` doi: 10.1145/3204949.3208125

Peis, I., López-Moríñigo, J.-D., Pérez-Rodríguez, M. M., Barrigón, M.-L., Ruiz-Gómez, M., Artés-Rodríguez, A., & Baca-García, E. (2020). Actigraphic recording of motor activity in depressed inpatients: a novel computational approach to prediction of clinical course and hospital discharge. *Scientific reports*, *10*(1), 17286.

Razavi, N., Horn, H., Koschorke, P., Hügli, S., Höfle, O., Müller, T., . . . Walther, S. (2011). Measuring motor activity in major depression: the association between the hamilton depression rating scale and actigraphy. *Psychiatry research*, *190*(2-3), 212–216.