HIGH SCHOOL EDITION
Journal of Student Research

Volume 11 Issue 4 (2022)

Diagnosing Major Depressive Disorder using Activity Data from Wearable Sensors and Machine Learning

Shlok Desai[1] and Brianna Marsh[#]

[1]Dougherty Valley High School, USA
[#]Advisor

ABSTRACT

Major Depressive Disorder (MDD), a mood disorder, is the most common psychological disorder. MDD manifests itself through a range of deadly symptoms, while diagnosis remains difficult and costly, often requiring psychiatrists or specialized techniques. An easier, and possibly early, diagnosis could improve treatment and outcomes. To address this unmet need, we developed a novel machine learning algorithm to detect MDD based on an individual's activity data i.e., movement combined with light data. The dataset from Kaggle.com included activity data for fifty-five participants in 2021. Our algorithm determined that disturbances in activity is a symptom that can be used to predict Major Depressive Disorder. This insight has the potential to accurately detect and diagnose a person with MDD. In conclusion, the algorithm connecting activity to MDD paves the way to an easier and effective way of diagnosing depression.

Introduction

According to Johns Hopkins Medicine (Mental Health Disorder Statistics, 2019), an estimated 26% of adult Americans (1 in 4 adults) suffer from a diagnosable mental disorder in a given year. The UN stated that depression is the leading cause of disability worldwide, and estimated that it affects more than 300 million people, or 4.4_% of the global population (UN Health Agency Reports Depression Now "Leading Cause of Disability Worldwide," 2017), and 13.3% of the population in the US alone (Kessler et al., 2003). This disorder could lead to permanent physical trauma, through self-harm and/or substance abuse, emotional damage, and even suicide. Major depressive disorder is a global and pertinent issue that affects every age group and country. The symptom we are focused on is a lack of activity experienced by a person with depression. The American Psychology Association (APA) attributes a lack of exercise as a prominent symptom of depression and poor mental health (Weir, 2011). However, diagnosing a person with depression is difficult without a professional, like a psychiatrist, which are also becoming harder to find. In fact, 111 million Americans live in mental health professional shortage areas (Weiner, 2018). So, our machine learning algorithm gives people a tool to measure if they or someone they know might have depression without the use of an expensive psychiatrist and it could be early. As a result, the question we attempted to answer is: can we predict Major Depressive Disorder through Feed Forward Neural Networks using one's activity data? We indeed answered this question and accurately predicted whether a person has Major Depressive Disorder or not with the data from a Depression Dataset on Kaggle.com (Mobius, 2021).

Definitions and Prior Work

It is important that we define several terms and concepts. Firstly, we must define the conditions we wish to diagnose using our algorithm. Major Depressive Disorder (MDD), or Depression, is a mood disorder characterized by symptoms such as, changes in appetite or weight, sleep, and/or psychomotor activity; decreased energy; feelings of worthlessness or guilt; difficulty thinking, concentrating, or making decisions; or recurrent thoughts of death, according to DSM-5 (American Psychiatric Association, 2013). It is the most common psychological disorder and is the main disorder people reference when talking about mental health. Our data contains activity data, specifically ActiGraph data. The activity data of each of the participants is contained in an excel file, and it is recorded using an ActiGraph watch that they must wear throughout the duration of the study. This watch uses light and movement to determine one's activity, circadian rhythm, and more (Actigraphy 2017).

Secondly, for the technical aspects of this project, we used a machine learning, feed forward algorithm. Machine learning (ML) is a field of computer science, in which an algorithm improves itself through more experience and data. Essentially, it is self-learning. A neural network is used in artificial intelligence and machine learning and consists of various, interconnected layers with neurons that are activated when a certain thing occurs, like neurons in our brains, hence the name neural network. It learns through training and then tests itself after it has been trained to increase it's accuracy. For our project we used K fold Validation which trains the model by using K-1/K of the dataset and tests on 1/K and alternates through the whole dataset. We further improved the training by adding an adaptive learning rate, which decays the learning rate of each parameter group by a predetermined factor, called "gamma", every epoch. There are also multiple parameters that affect the performance of the model including: “hidden_size” (the number of layers), the “learning_rate” (the step size at each iteration), “num_epochs” (the number of complete passes of the dataset through the algorithm), and the “batch_size” (the number of samples that will be propagated through the network). There are multiple types of neural networks, each with its own applications and complexities. However, the one we used is a feed forward network. In this neural network, the data is passed into the network and multiplied by the weights and an output is given, then the weights are updated after every batch. Additionally, our one does have hidden layers making it more complex and able to compute larger inputs. To implement the algorithm we used Python, a programming language, in accordance with libraries such as Numpy, Pandas, MatPlotLib, and PyTorch as our main package to analyze and create our algorithm in its entirety.

Prior Work

Other notebooks (Oseh, 2021) (X, 2021) on this dataset are strictly plotting the data, but (Oseh, 2021) also adds a SVM, or Support Vector Machine, that is an algorithm that creates a line which separates the data into sides (classes). They use a variety of libraries to create unique plots of different aspects of the data. These graphs showed an evident correlation between activity and mood disorders. Some notebooks, like (Oseh, 2021) and (X, 2021), used the ActiGraph data to create plots comparing the average activity of control and patient populations. Two of these plots are shown in Figure 1.

It was clear that the “control” (non-diagnosed participants) had visually distinct activity patterns from the diagnosed participants. However, other notebooks like (X, 2021), also create graphs of the MADRS data that helps to understand how the depressed population’s MADRS data changes depending on their activity and how the control and condition group differs in terms of MADRS scores. In addition, a summary of research done outside of these notebooks that uses predictive machine learning algorithms in order to diagnose Depression can be found in Table 1.

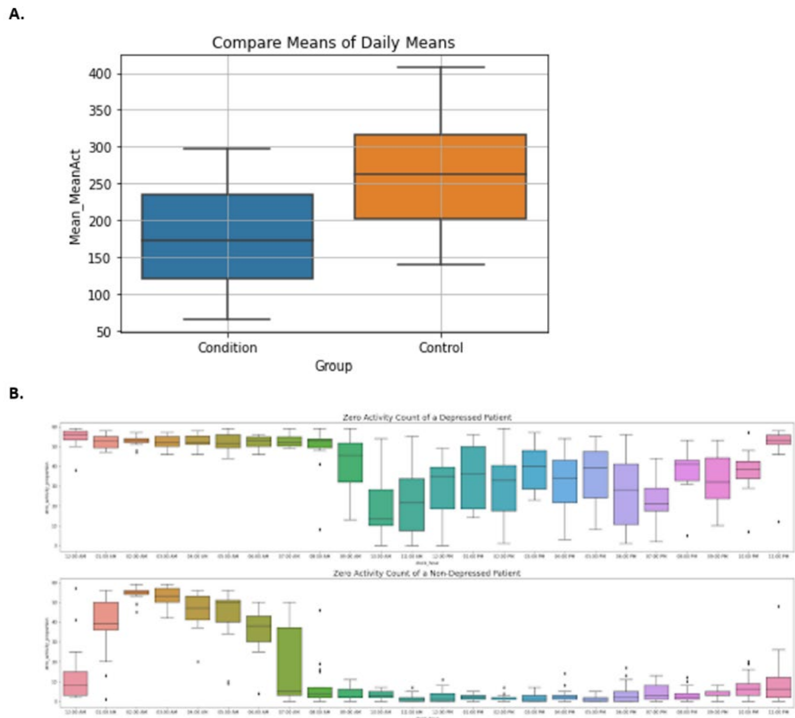

Figure 1. Plots from (Oseh, 2021) and (X, 2021). Figure 1A shows that the daily activity average for the control group is greater than that of the condition group (X, 2021). As seen in figure 1B, activity data is different every hour between the control and condition groups (Oseh, 2021).

Table 1. A Review of Prior Papers

Reference	Year Published	Algorithm	Methodology	Results	Limitations
(Haque et al., 2021)	2021	Random Forest Classifier (RF)	They used the second Australian Child and Adolescent Survey of Mental Health and Wellbeing 2013–14 survey data and used it in their Feed Forward Network	If a child/teen has 5 out of the 11 features in the dataset (including unhappiness, irritable mood, diminished interest, suicide attempt, etc. then they are at risk of having depression	This algorithm only predicts depression using basic questionnaire data. In addition, this dataset is only applicable for children and teenagers as they were the population from the dataset.
(Su et al., 2021)	2021	Random Forest Classifier (RF), Bagging, SVM, k-nearest neighbor	Their dataset was from the Chinese longitudinal healthy longevity study (CLHLS) which	They were able to use the data and a variety of algorithms to determine risk factors for	They use a variety of data and algorithms, however, their predictions are variable to

Reference	Year Published	Algorithm	Methodology	Results	Limitations
			includes health details and survey data of ~1500 Chinese elders. They then used 6 ML models to predict different depression risk factors and the depression risks in the population in the next two years	depression in elderly people (age 65 and older) in China, using the CLHLS dataset	change depending on the county of the person. Also, the questionnaire structure and the CLHLS waves meant LSTM model couldn't utilize more wave data in order to predict the risk factors of depression
(Cho et al., 2021)	2019	Random Forest Classifier (RF)	Their population was 55 patients with mood disorders for 2+ years. They used the self-reported moods and light exposure (through the sensor) by using a smartphone app. Then, from daily worn activity trackers they collected activity, sleep, and heart rate data. This was used to make a mood prediction algorithm.	They predicted the mood for each type of participant for the next 3 days. Also, the accuracies of all patients for no episode (NE), depressive episode (DE), manic episode (ME), and hypomanic episode (HME) were 85.3%, 87%, 94%, and 91.2%	Using self-reported data could lead to inaccurate data, however, it was probably diluted by the amount of authentic data from sensors. Also, more genetic assessments related to circadian rhythms were not utilized. Lastly, this algorithm would be harder to generalize.

Methods and Results

Our algorithm begins by loading all the relevant data from all fifty-six excel, or .csv, files that are provided by the dataset (Mobius, 2021). There are thirty-two control excel files (not diagnosed people), twenty-three condition files (people diagnosed with a mood disorder), and one "scores" excel file (includes demographic data of the participants). Each control and condition excel file has the following columns: timestamp, date, and activity. Additional "scores" excel file that contains demographic information like gender, age, employment, marital status, inpatient treatment status, and others. In addition, its file has the following columns for each participant: the number (patient identifier),

HIGH SCHOOL EDITION
Journal of Student Research

afftype (1: bipolar II, 2: depression, 3: bipolar I), melanch (1: melancholia, 2: no melancholia), madrs1 (MADRS score when measurement started), madrs2 (MADRS when measurement stopped). At first we calculated the mean of all the activity data of each participant then we calculated the log of the mean activity to make the skew (the measure of the asymmetry of the data around the mean) even less. We called it the “mean_log_activity”, and then we loaded all 55 rows of the participant number and mean_log_activity into 2 panda DataLoaders called “train_loader” and “test_loader” (for training and testing the model, respectively) with batch sizes of 64. We then used a for loop to run through train_loader the model and used another for loop to use test_loader to test the model. However, since the variation in the data was so small the model had a tough time predicting if a person has a condition or not. The model’s accuracy ranged from 50% to 55%. We soon figured out that the numbers were too similar for the model to differentiate and there were only fifty-five numbers in train_loader and test_loader combined. Thus, we changed the data that we used in the train and test loader to include the “mean_activity” (average activity of each participant over a time frame) instead of “mean_log_activity”, because it made sure the values were further apart. In addition, we got the mean of each participant for every hour, rather than their activity throughout the whole study. This added variance in the data for our model to utilize. The result was not so different; the accuracy was between 40% and 50%. Then we altered some of the parameters like the “hidden_size”, the “learning_rate”, “num_epochs”, and the “batch_size”. These changes were already helping our algorithm’s accuracy a lot, but it wasn't going too high. So, we decided to add more data. We added two columns which were the average standard deviation (the variance in the data) of all the data of every hour throughout the study, and the average of all the zero activity (how many times the activity was at 0) for every hour. We decided to use Matplotlib to plot this data and see other variables we could use (Figure 2).

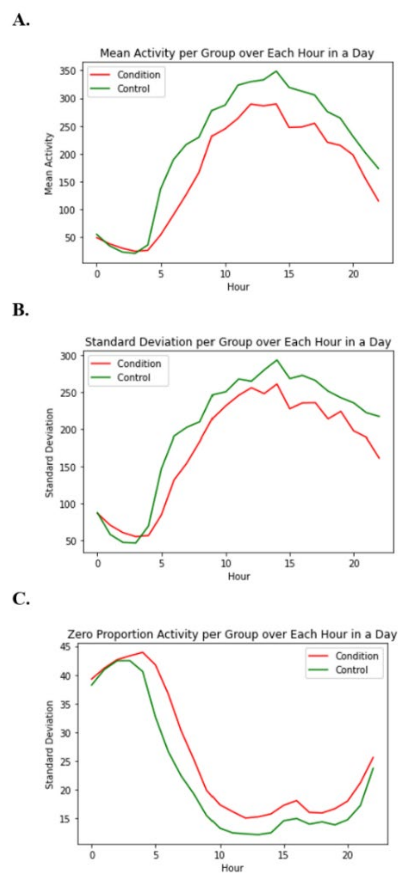

Figure 2. Matplotlib Plots. Figure 2A, 2B, and 2C are the plots we created that encouraged the use of mean activity, standard deviation, and zero proportion activity.

We added these features to our model because depressed people would have a higher zero count and lower standard deviation, and they intersect and switch which is a distinct feature. These 2 things would make it easier for the model to diagnose the participant. To improved our training, we implemented K fold validation, and an adaptive learning rate by using PyTorch's ExponentialLR. We used 4 folds in our K fold validation and had relatively high gamma to keep our model stable. When we used this expanded dataset and improved training and testing our model had testing accuracies of 68% to 73%, which demonstrated how model improved due to these changes. Our training and testing accuracy can also be seen in Figure 3 where we plotted the accuracy of the model while training and testing for each epoch.

A.

B.

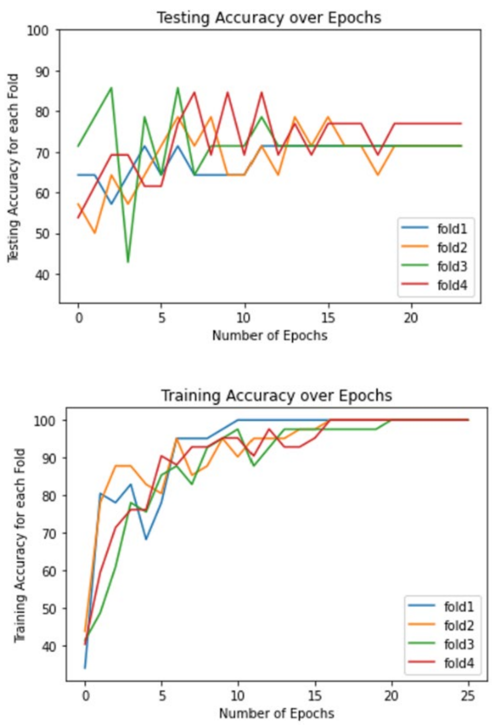

Figure 3. Sample Testing and Training Accuracies. Figure 3A how our model performed while testing while Figure 3B represents how our model performed during the training process.

Conclusion and Importance

In conclusion, our model demonstrates that activity data can be utilized to accurately diagnose the likelihood of MDD in a person, due to the fact that it repeatably averaged between 68\% and 73\%. Our algorithm was created using Python and it's many packages. We gathered the average mean, standard deviation, and zero activity count for every hour of the day for each person. Then we used K-fold validation in accordance with a dynamic learning rate to train and test our model. We also changed the parameters to best fit our model and the final values for each of these parameters can be found in Table 2.

Our model is important in providing an accessible and early tool in diagnosing the possibility of mood disorders like Major Depressive Disorder which affect many people and is the most common psychological disorder.

Table 2. Parameters used and values for each

Parameter	Value
hidden_size	200
num_epochs	24
batch_size	4
k (number of folds)	4
learning_rate	0.00075
Number of Layers	3
gamma	0.99

Future Directions

This model and dataset have a lot of future directions for a machine learning algorithm. In the future, we could use the activity data and the MADRS survey data to see the extent of how someone's depression has progressed/gotten worse. MADRS, or Montgomery–Åsberg Depression Rating Scale, is a widely used clinician-rated measure of depressive severity (Quilty et al., 2013). And, this dataset includes each person's MADRS score before the study starts and after it ends. Also, since people with bipolar have alternating periods of depression (low activity) and mania (hyper activity) (Mayo Clinic Staff, 2021) you could get the average of the activity data for every week and create a model that differentiates between someone who has MDD and bipolar disorder. In addition, combining other measures that are known symptoms of depression, such as weight/BMI change, hours slept, and others. This will help improve the accuracy of the model when trying to predict if someone has depression.

References

Actigraphy. Stanford Health Care (SHC) - Stanford Medical Center. (2017, September 12), from https://stanfordhealthcare.org/medical-tests/s/sleep-disorder-tests/procedures/actigraphy.html

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Cho, C. H., Lee, T., Kim, M. G., In, H. P., Kim, L., & Lee, H. J. (2019). Mood Prediction of Patients With Mood Disorders by Machine Learning Using Passive Digital Phenotypes Based on the Circadian Rhythm: Prospective Observational Cohort Study. Journal of medical Internet research, 21(4), e11029. https://doi.org/10.2196/11029

Haque UM, Kabir E, Khanam R (2021) Detection of child depression using machine learning methods. PLOS ONE 16(12): e0261131. https://doi.org/10.1371/journal.pone.0261131

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, A. J., Walters, E. E., & Wang, P. S. (2003). The Epidemiology of Major Depressive Disorder. JAMA, 289(23), 3095. https://doi.org/10.1001/jama.289.23.3095

Mayo Clinic Staff. (2021, October 29). Mood disorders. Mayo Clinic, from https://www.mayoclinic.org/diseases-conditions/mood-disorders/symptoms-causes/syc-20365057

Mental Health Disorder Statistics. (2019, November 19). Hopkins Medicine, from https://www.hopkinsmedicine.org/health/wellness-and-prevention/mental-health-disorder-statistics

Mobius. (2021). The Depression Dataset, Version 1, from https://www.kaggle.com/datasets/arashnic/the-depression-dataset

Oseh, S. (2021, August 3). Depression-dataset analysis and Machine Learning. Kaggle, from https://www.kaggle.com/code/samuelkali/depression-dataset-analysis-and-machine-learning/output

Quilty, L. C., Robinson, J. J., Rolland, J. P., Fruyt, F. D., Rouillon, F., & Bagby, R. M. (2013). The structure of the Montgomery-Åsberg depression rating scale over the course of treatment for depression. International journal of methods in psychiatric research, 22(3), 175–184. https://doi.org/10.1002/mpr.1388

Su, D., Zhang, X., He, K., & Chen, Y. (2021). Use of machine learning approach to predict depression in the elderly in China: A longitudinal study. Journal of affective disorders, 282, 289–298. https://doi.org/10.1016/j.jad.2020.12.160

UN health agency reports depression now “leading cause of disability worldwide.” (2017, February 23). UN News, from https://news.un.org/en/story/2017/02/552062-un-health-agency-reports-depression-now-leading-cause-disability-worldwide#:~:text=Depression%20is%20the%20leading%20cause,young%20people%20and%20the%20elderly.

Weiner, S. (2018, February 12). Addressing the escalating psychiatrist shortage. AAMC, from http://www.aamc.org/news-insights/addressing-escalating-psychiatrist-shortage

Weir, K. (2011, December). The exercise effect. American Psychological Association, from https://www.apa.org/monitor/2011/12/exercise

X, C. (2021, April 8). Depression and motor activity. Kaggle, from https://www.kaggle.com/code/docxian/depression-and-motor-activity/comments