Actigraphy Based Depression Analysis

Umeshraja N 22m1078[1], Dhruv Kudale 22m2116[1], Ameya Srivastava 22m2118[1], and Kishan Maharaj 22m2122[1]

[1]Indian Institute of Technology Bombay

April 1, 2023

Abstract

Depression is a mood disorder with a persistent feeling of sadness that can affect the patient physically and emotionally which may cause disturbances in normal day-to-day life. Even though effective long-term treatments are available, an early diagnosis is always beneficial. Measuring motor activity could be one of the possible ways to provide an early warning for the same. We thoroughly discuss the approach of detecting depression from gross motor activity data.

1 Introduction

More than 280 million individuals of all ages (or around 3.5% of the world's population) suffer from depression, a mental condition characterized by low mood and aversion to action.[9] Depression has an impact on a person's motivation, thoughts, behavior, feelings, and sense of well-being.[9] According to the World Mental Health Atlas 2014, there are 0.30 psychiatrists for every 100,000 people in India.[6] These facts suggest a very strong requirement of automated depression detection in an affordable way.

We suggest an actigraphy-based approach that can be easily integrated with modern smartwatches and can help in the early diagnosis of depression. This method can not only help in detecting depression but also in preventing it.

The following sections describes the data and the related work that has been carried out in this field. Later sections explains our methodology and implementation details along with the results. The corresponding code is open source and can be found on GitHub.[1].

2 Data Description and Related Work

We have considered the "Depresjon" dataset introduced in [5] which contains the motor activity of individuals (clinically depressed and healthy) recorded per minute for several days. The total number of individuals involved is 55 which is composed of 23 depressed subjects and 32 healthy subjects. The motor activity is determined through continuous monitoring by means of actigraphy. The severity of the patient's depressive state is labeled by medical experts through Montgomery-Asberg Depression Rating Scale (MADRS).

This dataset is suitable for comparing different machine learning classification approaches such as feature-based or deep learning-based methods for time series. Several approaches like [1] have been studied in time series classification tasks. [5] discusses about the usage of raw time series of length 1440 and passing it to classical machine learning models like Nearest Neighbors, Support Vector Machine (SVM), Decision Tree, Random Forest, etc. to classify the series into Depressed or non Depressed classes. Convolutional

[1]https://github.com/kishanmaharaj/Actigraphy-based-Depression-Analysis

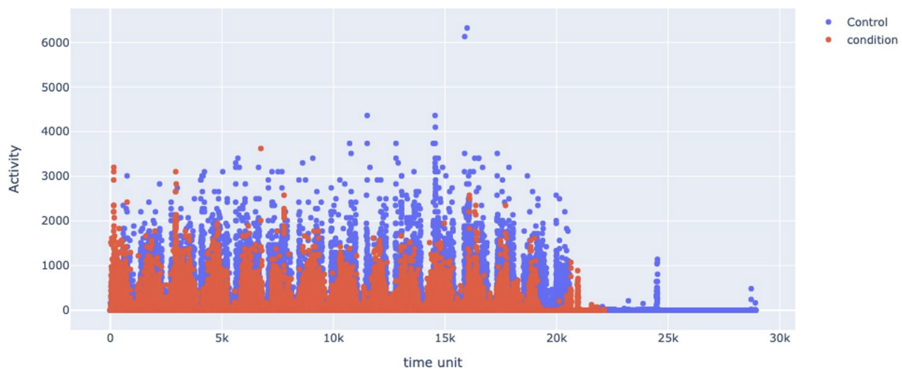

Figure 1: Activity comparison for Control (healthy) and condition (Depressed)

Neural Networks have also been employed to deal with time series classification. [4] discusses about implementing a Convolutional Neural Network to perform classification tasks without explicit feature extraction.

Figure 1 shows the scatter plot of Control and Condition group. It is evident that gross motor activity of Condition group is significantly less as compared to Control group. These facts aligns with the results of psycho-motor studies done in psychology.[8]

3 Methodology

We have explored both the Machine Learning and Deep Learning paradigms for this problem of time series classification. Feature extraction and feature selection was used in order to proceed with Machine Learning models. Deep Learning based approaches were considered because of their abilities to eliminate the requirements of feature engineering, which is a very time consuming and tedious task.

3.1 Ensemble Classifiers

We were initially exploring approaches that used Machine Learning models for tackling such problems for which we had performed feature engineering. Highly interpretable ML models which include Linear Regression and Decision trees were also included in our experimentation. Apart from classification, we were thinking along the lines of understanding the most important features for determining the target. During the stage of exploratory data analysis, we found out that not only there is significant overlap in both classes but also a considerable presence of outliers. This drew our attention to ensemble classification. Since we already had used Random Forests for extracting top features, we decided to proceed in that direction.

The objective of using **Random Forest Classifiers** was to reduce variance and grant all features an equal chance to exhibit their importance. After having a look at the considerably promising results of Random Forest Classification we decided to use another ensemble classifier that relies on gradient boosting mechanism. We have thus incorporated **XGBoost Classifier** for enhancing the overall predictive performance. Figure 3.1 shows the basic ensemble classifier architecture like XGBoost [3]. The weak learners in Random forest classifiers are decision trees with randomly selected attributes whereas in the case of XGBoost, the weak learners are gradient-boosted trees.

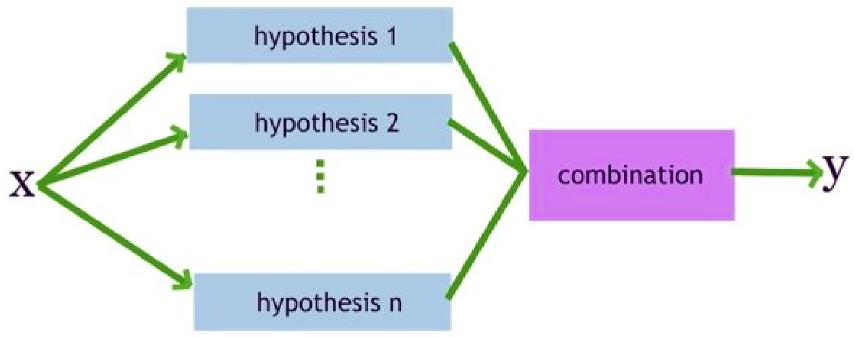

Figure 2: A general ensemble architecture

3.2 LSTM

Recent studies [7] have highlighted the advantages of leveraging LSTM and its variants for time series classification tasks. We have considered Vanilla LSTM for this task. This approach not only eliminates the need for explicit feature extraction/selection but also handles time-variant data well which is beneficial for this problem since the data contains temporal features. LSTMs are proven to be resistant to noise which will be advantageous for certain cases where subjects exhibits fluctuating or outlying motor activity patterns.

3.3 Incorporating Oversampling

The Depresjon dataset [5] contains around 34.8 % of the Condition class, which makes this a case of moderately unbalanced data. This might add bias towards the majority class. This suggests a need for oversampling the minority class.

We have incorporated SMOTE[2], a technique that oversamples the minority class by using "synthetic" data points, instead of oversampling with replacement. SMOTE takes each minority class sample and inserts synthetic samples along the line segments joining any of the k minority class nearest neighbours and the selected data point (of minority class). Randomly selected neighbours from the k nearest neighbours are dependent on the quantity of over-sampling needed.

4 Implementation

4.1 Data Processing

This step involves converting the raw time series data into day-wise gross motor activity. For this, we have partitioned the raw series of all subjects for the entire duration considered into fixed-length sequences of dimension 1440 each. As a result of this, we retrieve around a thousand records comprising 670 Control class samples and the rest of the samples belonging to the Condition class.

4.2 Feature Extraction

This step involves extracting the numerical data from the raw dataset without the loss of original information. Raw data would be a collection of time series data and the features that are extracted are based on spectral, temporal, and statistical features. 60 features were extracted using Time Series Feature Extraction Library (TSFEL). The pipeline involved in TSFEL is been shown in Figure 2.

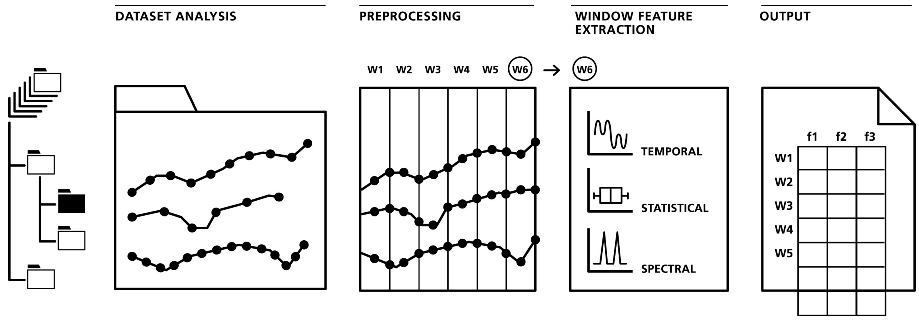

Figure 2: Steps involved in feature extraction

4.3 Feature Selection

We have used Random Forest for selecting the top 10 features. The goal is to quantify the drop in data accuracy when the values for that feature are randomly permuted. The feature is irrelevant if the drop is small, and vice versa. The following are the features which were selected based on the importance score which is measured as an average impurity decrease or "Gini Importance". Each feature represents the characteristics of a Condition and Control subjects.

- Total Energy - Sum of the squares of the Motor Activity
- Spectral roll-on - Frequency at which most of the energy is contained below this value.
- Positive turning points - Temporal feature which gives the number of positive turning points
- Mean - Statistical average value of the signal
- Neighbourhood peaks - Number of peaks from a certain point in the time domain
- Spectral positive turning points - Number of times the response in the frequency domain turns positive
- Negative turning points - Temporal feature which gives the number of negative turning points
- Signal distance - Total distance joining all points in the signal
- Interval distribution - Counts of all values in the signal
- Zero crossing rate - Number of times Signal turns from positive to negative and vice versa.

4.4 Training and Testing

After completing the feature extraction, we have used the top 10 features to train ensemble classifiers namely the Random forest classifier and XGBoost classifier. We have also trained these models after performing oversampling using SMOTE.

4.4.1 Training Details for Ensemble Models

The following open-source libraries and tools were used for training ML models:

1. Numpy: for mathematical operations
2. Pandas: handling data Frame

3. Plotly: for Data Visualizations in exploratory data analysis

4. Sklearn: for Random Forest classification

5. Xgboost: for XGboost classifier

Furthermore, we have trained an LSTM model which takes a raw input sequence of length 1440 to predict the concerned target value. Considering the size of our dataset, we have tested all our models using ten-fold cross-validation.

4.4.2 Training Details for LSTM Model

We have used the Keras library to train an LSTM network on the following hyperparameters:

1. Learning Rate: 0.001

2. No of Epochs: 50

3. Hidden Layers: 4

4. Activation Function: ReLu

5. Units per layer Layer: 256

6. Sequence Length: 1440

5 Results and Discussion

Our results are tabulated in Table 1. We compare our results of the Random forest Classifier and XGBoost Classifier (both with and without SMOTE) with baseline models which uses Linear SVMs [5] and one-dimensional CNN's [4]. The best performance is achieved by leveraging the XGBoost Classifier with oversampling. Figure 5 displays the confusion matrix corresponding to testing carried out using the same. Figure 3 displays the performance of a Random forest classifier without oversampling. A significant number of misclassifications of depressed class to not-depressed class is observed. This is possibly due to the class imbalance resulting from the presence of only 34.8 percent of records corresponding to the depressed class. The accuracy increases when accompanied by oversampling. Figure 4 highlights the performances of Random Forest Based classifiers with SMOTE.

Apart from that, we have also trained an LSTM model for the classification task, the results for which are displayed in the confusion matrix of Figure 6. The possible reason for the comparatively poor performance of the LSTM model is its inability to properly figure out the relevant feature set from the limited amount of training data. The considerably important features have been explicitly provided for machine learning models which justify their better performances. Additionally, the length of the input sequence provided to this LSTM model is sufficiently large (of size 1440) which proves to be a challenging task to handle.

Also, there exists a considerable portion of outlying individuals who are being misclassified. This is analogous to the fact that depressed individuals can exhibit gross motor activity that can match their non-depressed counterparts. Similarly, a small population of non-depressed individuals exhibited comparatively less gross motor activity and were misclassified as depressed individuals. This further highlights the inherent class overlap problem which is present in the collected data.

Model	Precision	Recall	Accuracy	F1-Score
XGBoost with SMOTE	**0.8459**	**0.8828**	**0.8604**	**0.8628**
Random Forest with SMOTE	0.8275	0.8694	0.8440	0.8467
Random forest	0.7709	0.7512	0.7872	0.7557
XGBoost	0.7685	0.7568	0.7892	0.7594
Linear SVM [5]	0.735	0.729	0.727	0.724
LSTM	0.71	0.71	0.71	0.71
1D CNN[4]	0.71	0.70	0.71	0.70
Trivial Classifier	0.52	0.35	0.49	0.42

Table 1: Our Results

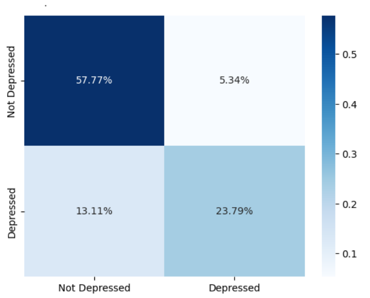

Figure 3: Results for Random Forest

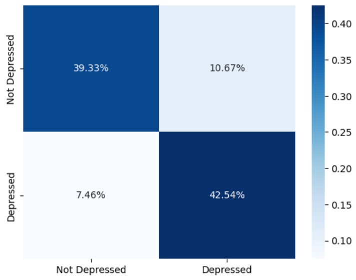

Figure 4: Results for RFC with SMOTE

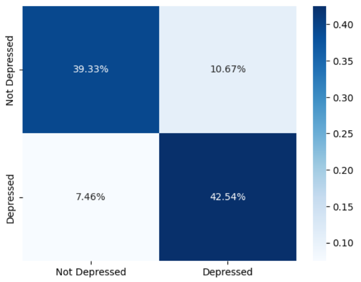

Figure 5: Results for XGB SMOTE

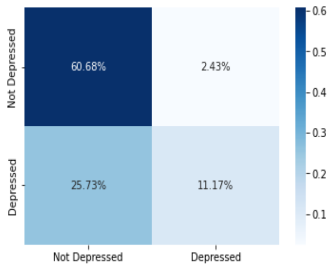

Figure 6: Results for LSTM

6 Conclusion

The task of depression detection can be simplified to the analysis of the daily gross motor activity of an individual which can alleviate the challenges encountered in classical psychiatric evaluation. This can also help countries like India to fulfill the scarcity of psychologists.

7 Future Work

One of the upcoming novel tasks in this domain is to possibly interpret the existing models and understand how each feature is contributing to diagnosing depression based on which insightful decisions can be taken. Efforts could be made in the direction of prevention of depression by continuous monitoring of gross motor activity. A threshold of daily gross motor activity can be determined which aligns with the decision boundary of the two classes. Proper planning and consistent maintenance of individual gross motor activity can prove to be beneficial in preventing depression.

References

[1] Amin Aminifar, Fazle Rabbi, Violet Ka I Pun, and Yngve Lamo. Monitoring motor activity data for detecting patients' depression using data augmentation and privacy-preserving distributed learning. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 2163–2169, 2021.

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, jun 2002.

[3] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[4] Joakim Ihle Frogner, Farzan Majeed Noori, Pål Halvorsen, Steven Alexander Hicks, Enrique Garcia-Ceja, Jim Torresen, and Michael Alexander Riegler. One-dimensional convolutional neural networks on motor activity measurements in detection of depression. In *Proceedings of the 4th International Workshop on Multimedia for Personal Health amp; Health Care*, HealthMedia '19, page 9–15, New York, NY, USA, 2019. Association for Computing Machinery.

[5] Enrique Garcia-Ceja, Michael Riegler, Petter Jakobsen, Jim Tørresen, Tine Nordgreen, Ketil J. Oedegaard, and Ole Bernt Fasmer. Depresjon: A motor activity database of depression episodes in unipolar and bipolar patients. In *Proceedings of the 9th ACM Multimedia Systems Conference*, MMSys '18, page 472–477, New York, NY, USA, 2018. Association for Computing Machinery.

[6] Kabir Garg, C Naveen Kumar, and Prabha S Chandra. Number of psychiatrists in india: Baby steps forward, but a long way to go. *Indian journal of psychiatry*, 61(1):104, 2019.

[7] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Samuel Harford. Multivariate lstm-fcns for time series classification. *Neural Networks*, 116:237–245, 2019.

[8] Christina Sobin and Harold A Sackeim. Psychomotor symptoms of depression. *American Journal of Psychiatry*, 154(1):4–17, 1997.

[9] Wikipedia contributors. Depression (mood) — Wikipedia, the free encyclopedia, 2022. [Online; accessed 26-November-2022].