Healthcare Transformation with Informatics and Artificial Intelligence
J. Mantas et al. (Eds.)

doi:10.3233/SHTI230529

Performance of Artificial Intelligence in Predicting Future Depression Levels

Sarah AZIZ[a,1], Rawan ALSAAD[a], Alaa ABD-ALRAZAQ[a], Arfan AHMED[a] and Javaid SHEIKH[a]
[a] *AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar*

Abstract. Depression is a prevalent mental condition that is challenging to diagnose using conventional techniques. Using machine learning and deep learning models with motor activity data, wearable AI technology has shown promise in reliably and effectively identifying or predicting depression. In this work, we aim to examine the performance of simple linear and non-linear models in the prediction of depression levels. We compared eight linear and non-linear models (Ridge, ElasticNet, Lasso, Random Forest, Gradient boosting, Decision trees, Support vector machines, and Multilayer perceptron) for the task of predicting depression scores over a period using physiological features, motor activity data, and MADRAS scores. For the experimental evaluation, we used the Depresjon dataset which contains the motor activity data of depressed and non-depressed participants. According to our findings, simple linear and non-linear models may effectively estimate depression scores for depressed people without the need for complex models. This opens the door for the development of more effective and impartial techniques for identifying depression and treating/preventing it using commonly used, widely accessible wearable technology.

Keywords. Depression, Artificial Intelligence, Motor Activity, Machine Learning, Depresjon.

1. Introduction

Depression is a prevalent mental health condition that affects around 280 million individuals worldwide [1]. It causes feelings of sadness, loss of interest in previously enjoyable activities, and can have negative impacts on one's emotional and physical well-being [2]. Early detection of depression is crucial to providing effective treatment and improving the quality of life for affected individuals. Therefore, wearable devices and artificial intelligence (AI) have been utilized together to detect and predict depression by collecting and analyzing various biomarkers or biosignals such as heart rate, physical activities, sleep patterns, blood oxygen, and respiration rate. This combination of wearable devices and AI, which is also called wearable AI, can overcome the challenges of current approaches in depression assessment, which are time-consuming, subjective, and challenging to repeat. The Depresjon [3] dataset has been frequently used as a resource in studies to evaluate the efficacy of wearable technology in depressive disorder prediction [4-10]. In previous research, deep learning and advanced machine learning models have primarily been used to analyze motor activity measurements to distinguish between depressed and non-depressed people. The main tool for assessing patients' levels

[1] Corresponding Author: Sarah Aziz, E-mail: saa4038@qatar-med.cornell.edu.

of depression is the Montgomery-Asberg Rating Scale (MADRS) in depresjon dataset. In order to determine the severity of depression, the MADRS involves a discussion with the patient and the observation of 10 items that are pertinent to depression. A system that focuses on examining biological signals produced by the body is necessary because depression scores determined through discussions with patients can be arbitrary. For the purpose of helping clinicians diagnose and treat depression, such a system could offer more unbiased information. In this study, we hypothesized that simple linear and nonlinear models, based on physiological characteristics, motor activity data, and MADRAS scores, could accurately predict depression scores in the future for depressed people.

2. Materials and Methods

The methodology comprised of a comprehensive depression dataset that was obtained using a wearable wristband was used for the predictive analysis of depression scores after going through a few feature engineering steps. The prediction results were validated by computing errors using the metrics RMSE and MAE.

The Depresjon [3] dataset, which contains information on patients whose motor activity was monitored by an actigraph watch worn on the right wrist, was used in this study. The database includes information on both the controls (32 participants without depression) and the conditions (presence of depression, 23 subjects). For this study only data corresponding to depressed participants for consider. Actigraph data over time and MADRS scores were the two groups into which the characteristics gathered for each participant were separated. The data collected includes features such as "timestamp", "date", and "activity" from the actigraph watch. The MADRS scores include features such as patient demographic and clinical information along with two MADRAS scores. It has two components that make up the MADRS score: MADRS-1 and MADRS-2. With scores ranging from 0 to 6, the MADRS-1 evaluates the severity of 10 distinct symptoms of depression. A higher score denotes more severe symptoms. By measuring the difference in symptom severity between the baseline and follow-up assessments, the MADRS-2, on the other hand, assesses the response to treatment, with negative scores indicating improvement and positive scores indicating worsening of symptoms.

For this study, only the wearable data (motor activity) characteristics of condition (depressed) subjects were used, along with each patient's unique features and MADRAS scores. While MADRAS-2 served as the dependent variable, MADRAS-1 served as the independent variable.

3. Results

Metrics like RMSE (Root Mean Square Error) and MAE(Mean Absolute Error), which penalize larger errors while emphasizing the average deviation, are used to assess the precision of predictive models. Both are popular and simple to understand, which makes them useful for assessing a model's performance and pinpointing its weak points. Figure 1 demonstrate the performance of each model in predicting depression level as measured by RMSE and MAE, respectively. As shown in these figures, all models have high and similar performance (RMSE: 2.32-3.56, MAE:1.89-2.97) in predicting depression level.

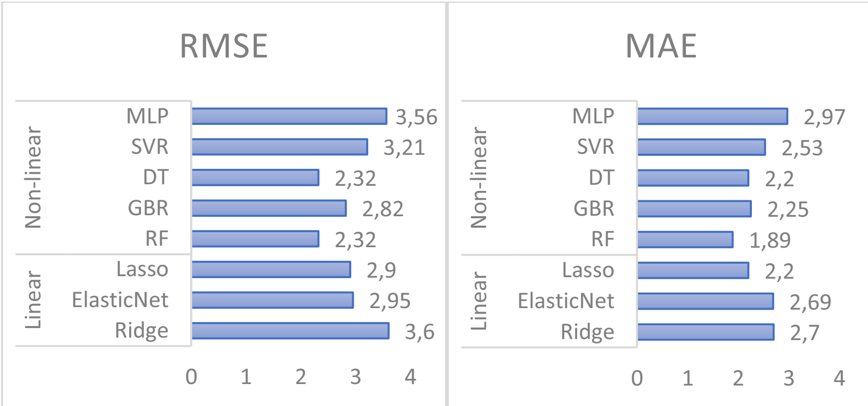

Figure 1. Performance of each model in predicting depression level as measured by RMSE (left) and by MAE (right)

4. Discussion and Conclutions

All regressors show their greatest performance for the selected characteristics, with not much variation in predictions between linear and non-linear models. Lasso (RMSE:2.9, MAE:2.2), a linear model, performed the best among the group. The most effective were tree-based regressors Decision Tree, Random Forest, and Gradient Boosting Regressor from non-linear models; one explanation for this can be their ability to perfectly capture the non-linearity among the features and target value, along with robust to outliers. RF and GBR being ensemble methods reduce the overfitting threat.Although applying ML models to this dataset produced promising results for estimating depression score from WD data, we report RMSE values of 2.32 and MAE values of 1.89 using ML algorithms, outperforming previous studies' complex models. This contrasts with previous studies, which reported RMSE values of 2.8 - 5.07 mainly using Deep Neural Network algorithms [7-12] as shown in table 1. Even though all of the models used in this study performed well, it is advised that linear models be prioritized before moving on to more intricate deep learning models. Better resource and energy management might result from this strategy. If more research is done to confirm Lasso's effectiveness using a bigger dataset, it might be able to outperform the other linear models in terms of benefits.

Table 1. Comparison of results with previous studies

Study	Model	Dataset used	RMSE	MAE
STD-1 [7]	CNN	Depresjon	5.60	-
STD-2 [9]	AdaBoost	closed	4.6	-
STD-3 [10]	XGBoost	closed	2.83	2.39
STD-4 [11]	Adaboost	closed	5.43	4.27
STD-5 [12]	XGBoost	closed	3.10	2.54
Our Study	RF	Depresjon	**2.32**	**1.89**
Our Study	Lasso	Depresjon	**2.9**	**2.2**

Lasso is renowned for its automatic feature selection ability. In addition to having research implications for the development of more effective and objective methods for diagnosing and treating depression, the paper's prediction of depression levels from wearable device data also has practical implications for the development of new tools for managing mental health and the potential for more individualized and successful interventions.

In this paper, we examined the performance of simple linear and non-linear algorithms for predicting depression scores using physiological features, motor activity data, and MADRAS scores. Model complexity is a key consideration in machine learning. Models with a high degree of complexity may be able to capture more variations in the data but are also more difficult to train and may be more prone to overfitting. Therefore, our results suggest that simple linear models can objectively detect depressed subjects without the need for complex models.

References

[1] Institute of Health Metrics and Evaluation. Global Health Data Exchange (GHDx).2021; Available from:http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/d780dffbe8a381b25e1416884959e88b.
[2] American Psychiatric Association. What Is Depression? 2022; Available from: https://www.psychiatry.org/patients-families/depression/what-is-depression#section_2.
[3] Garcia-Ceja E, Riegler M, Jakobsen P, Tørresen J, Nordgreen T, Oedegaard KJ, Fasmer OB. Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients. InProceedings of the 9th ACM multimedia systems conference 2018 Jun 12 (pp. 472-477).
[4] Choi J, Lee S, Kim S, Kim D, Kim H. Depressed mood prediction of elderly people with a wearable band. Sensors. 2022 May 31;22(11):4174.
[5] Dai R, Kannampallil T, Zhang J, Lv N, Ma J, Lu C. Multi-task learning for randomized controlled trials: a case study on predicting depression with wearable data. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2022 Jul 7;6(2):1-23.
[6] Espino-Salinas CH, Galván-Tejada CE, Luna-García H, Gamboa-Rosales H, Celaya-Padilla JM, Zanella-Calzada LA, Tejada JI. Two-Dimensional Convolutional Neural Network for Depression Episodes Detection in Real Time Using Motor Activity Time Series of Depresjon Dataset. Bioengineering. 2022 Sep 9;9(9):458.
[7] Frogner JI, Noori FM, Halvorsen P, Hicks SA, Garcia-Ceja E, Torresen J, Riegler MA. One-dimensional convolutional neural networks on motor activity measurements in detection of depression. InProceedings of the 4th International Workshop on Multimedia for Personal Health & Health Care 2019 Oct 15 (pp. 9-15).
[8] Galván-Tejada CE, Zanella-Calzada LA, Gamboa-Rosales H, Galván-Tejada JI, Chávez-Lamas NM, Gracia-Cortés M, Magallanes-Quintanar R, Celaya-Padilla JM. Depression episodes detection in unipolar and bipolar patients: A methodology with feature extraction and feature selection with genetic algorithms using activity motion signal as information source. Mobile Information Systems. 2019 Apr 23;2019.
[9] Ghandeharioun A, Fedor S, Sangermano L, Ionescu D, Alpert J, Dale C, Sontag D, Picard R. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In2017 seventh international conference on affective computing and intelligent interaction (ACII) 2017 Oct 23 (pp. 325-332). IEEE.
[10] Mullick T, Radovic A, Shaaban S, Doryab A. Predicting Depression in Adolescents Using Mobile and Wearable Sensors: Multimodal Machine Learning–Based Exploratory Study. JMIR Formative Research. 2022 Jun 24;6(6):e35807.
[11] Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhathena D, Fisher LB, Cusin C, Nyer M, Yeung A, Sangermano L. Monitoring changes in depression severity using wearable and mobile sensors. Frontiers in psychiatry. 2020 Dec 18;11:584711.
[12] Rykov Y, Thach TQ, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. JMIR mHealth and uHealth. 2021 Oct 25;9(10):e24872.