IntechOpen Journals

Digital Medicine and Healthcare Technology

NEWS & VIEWS

The Importance of Using Binary Classification Models in Predicting Depression from a Machine Learning Perspective

Soumya Choudhary* and Girish Srinivasan

Behavidence, Inc., New York, NY, USA
*Corresponding author. E-mail: soumya@behavidence.com

Citation
Soumya Choudhary and Girish Srinivasan (2022), The Importance of Using Binary Classification Models in Predicting Depression from a Machine Learning Perspective. *Digital Medicine and Healthcare Technology* 2022(0), 1–4.

DOI
https://doi.org/10.5772/dmht.12

Published
23 December 2022

1. News

Digital phenotyping (DP) has shown promise for personal mobile devices to be used for mental health assessment. DP relies on the concept of ecological momentary assessment (EMA) that involves repeated sampling of an individual's behaviors and experiences in real-time, in the person's natural environment. Kamath *et al.* [1], discuss the use of EMA, both active and passive, for evaluation of 'objective symptoms with subjective patient experiences'. This process involves training models with data extracted from smartphones (passive EMA) using standardized mental health assessment questionnaires (active EMA) as ground truth. Although these questionnaires serve as the best indicator to assess human behavior, they are accompanied with biases. Using this as ground truth creates challenges particularly in severity determination through phenotyping models. DP suits well for highly sensitive identification of problematic mental health behaviors through binary classifications of normal versus severe symptomatic behavioral presentation.

2. Views

The Diagnostic and Statistical Manual of Mental Disorders (5th edition; [DSM-5]) categorizes psychiatric symptomatology into specific disorders. Despite evidence supporting such categorization, diagnosis remains subjective, as self-reporting by patients remains the foundation of clinical evaluation. The 9-item Patient Health Questionnaire (PHQ-9) and the 7-item Generalized Anxiety Disorder questionnaire (GAD-7) are commonly used patient self-rated instruments for depression and anxiety respectively. PHQ-9 is evidenced to have inconsistencies in its cut-off, understanding, and application [2]. A study done on usage of PHQ-9 in clinics revealed that there was significant variability in the interpretation of the questions, responses and scores across clinicians and patients [2]. The GAD-7 has been shown to not discriminate well in the lower spectrum of anxiety [3] suggesting its applications are restricted to severe grade anxiety disorders.

Technology now enables an accurate and holistic measurement of patients' lived experiences. DP is a new and exciting field that analyzes passive data from a user's smartphone (screen duration, number of locks/unlocks, sensor data, etc) using advanced analytics such as machine learning to develop a digital behavior profile for the user [1]. This digital profiling shows promise to be used for mental health assessments and screening, so that interventions can be provided effectively and at the right time.

The challenge with DP comes when machine learning models use standardized questionnaires as "ground truth" to classify users in mild, moderate and severe diagnostic groups (multi-class classification) as compared to none versus severe group (binary classification). First, there is bias-creep due to factors such as under or over reported symptoms, different understanding of the questions and most importantly different cut-off ranges for different cultures [4]. Second, the overlapping behavioral patterns in the intermediate groups of mild, moderate, and moderately severe categories create further confusion, as scales designed for screening are used for severity diagnosis [5]. This makes finding features by machine learning algorithms that pick-up small significant changes in user behavior more challenging. Third, inherent imbalance in the data across different classes induces exacerbated data bias and overfitting. The comparisons between multi-class and binary machine learning models discussed above have been evidenced in published findings. Nguyen *et al.* demonstrated the use of ML models (such as Support Vector Machine) to predict severity of anxiety based on GAD-7 and the results showed an accuracy of 94%–98% to classify between minimal vs severe scores. When analyzing minimal and mild versus moderate and severe using a GAD-7 score of 10 as a cut-off, the accuracy dropped to 87%–92%. The 4-class classification model achieved only 64%–74% accuracy [6]. Yue *et al.* used internet traffic characteristics to classify whether one has depression or not using machine learning models and achieved an accuracy of 80% [7]. Asare *et al.* used Random Forest binary classification models to classify the passive behavior into depressed v/s nondepressed groups with accuracy levels of 96–98% [8]. Additionally, a novel app-based solution demonstrated an accuracy of 87% and 76% for detecting behaviors similar to severe depressive disorder [9] and severe anxiety disorder [10] respectively, using only non-private smartphone usage data.

3. Closing remarks

Abundance of features engineered from passive data collection on mobile devices, enables bias free DP with zero respondent burden. Lack of consistency in questionnaire-based ground truth limits the application of such phenotypes to binary differentiation between presence versus absence of a condition. To apply DP powered by machine learning techniques for multi-class mental health severity determination, we need large amounts of clean balanced data for training or unsupervised data clustering methods such as self-organizing-maps or K-means clustering. Further, longitudinal studies on binary classification outcomes are needed to explore the possibility of using confidence metrics reported from these models as a mechanism to perform severity grading.

Conflict of interest

Authors SC and GS have jointly worked in developing the Behavidence App and are now employed at the company.

Acknowledgements

We would like to thank Roy Cohen, Dr Janine Ellenberger for their support in reviewing this article.

References

1 Kamath J, Leon Barriera R, Jain N, Keisari E, Wang B. Digital phenotyping in depression diagnostics: Integrating psychiatric and engineering perspectives. *World J Psychiatry*. 2022;**12**(3):393–409.

2 Ford J, Thomas F, Byng R, McCabe R. Use of the Patient Health Questionnaire (PHQ-9) in Practice: Interactions between patients and physicians. *Qual Health Res*. 2020 Nov;**30**(13):2146–2159.

3 Jordan P, Shedden-Mora MC, Löwe B. Psychometric analysis of the Generalized Anxiety Disorder scale (GAD-7) in primary care using modern item response theory. *PLoS One*. 2017 Aug;**12**(8):e0182162. doi:10.1371/journal.pone.0182162. PMID: 28771530; PMCID: PMC5542568.

4 Zimmerman M, Morgan TA, Stanton K. The severity of psychiatric disorders. *World Psychiatry*. 2018;**17**: 258–275. doi:10.1002/wps.20569.

5 Wittkampf K, van Ravesteijn H, Baas K, van de Hoogen H, Schene A, Bindels P, et al. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry*. 2009 Sep–Oct;**31**(5):451–459. doi:10.1016/j.genhosppsych.2009.06.001. Epub 2009 Jul 10. PMID: 19703639.

6 Nguyen B, Ivanov M, Bhat V, Krishnan S. Digital phenotyping for classification of anxiety severity during COVID-19. *Front Digit Health*. 2022 Oct;**4**: 877762. doi: 10.3389/fdgth.2022.877762. PMID: 36310921; PMCID: PMC9612961.

7 Yue C, Ware S, Morillo R, Lu J, Shang C, Bi J, et al. Automatic depression prediction using internet traffic characteristics on smartphones. *Smart Health (Amst)*. 2020 Nov;**18**: 100137. doi: 10.1016/j.smhl.2020.100137. Epub 2020 Sep 8. PMID: 33043105; PMCID: PMC7544007.

8 Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR Mhealth Uhealth*. 2021 Jul;**9**(7):e26540. doi:10.2196/26540. PMID: 34255713; PMCID: PMC8314163.

9 Choudhary S, Thomas N, Ellenberger J, Srinivasan G, Cohen R. A machine learning approach for detecting digital behavioral patterns of depression using nonintrusive smartphone data (Complementary Path to Patient Health Questionnaire-9 Assessment): prospective observational study. *JMIR Form Res*. 2022 May;**6**(5):e37736. doi:10.2196/37736. PMID: 35420993; PMCID: PMC9152726.

10 Choudhary S, Thomas N, Alshamrani S, Srinivasan G, Ellenberger J, Nawaz U, et al. A machine learning approach for continuous mining of nonidentifiable smartphone data to create a novel digital biomarker detecting generalized anxiety disorder: prospective cohort study. *JMIR Med Inform*. 2022 Aug;**10**(8):e38943. doi:10.2196/38943. PMID: 36040777; PMCID: PMC9472035.