Machine learning assessment of risk factors for depression in later adulthood

Fengqing Zhang,[a]* and Jiangtao Gou[b]

[a]Department of Psychological and Brain Sciences, Drexel University, 3201 Chestnut Street, Philadelphia PA 19104, USA
[b]Department of Mathematics and Statistics, Villanova University, 800 E. Lancaster Ave. Villanova, PA 19085, USA

The Lancet Regional Health - Europe 2022;18: 100399
Published online xxx
https://doi.org/10.1016/j.lanepe.2022.100399

Depression is the most prevalent mental health problem in older adults and poses substantial public health and economic burden. Investigation of the underlying risk factors of depression may improve the identification of at-risk individuals and guide future efforts for treatment of depression. Recent studies have found that lack of social support, poor health, and mobility are important risk factors for depression among older adults from the United Kingdom.[1,2] However, it is unclear whether these findings can be generalised to other populations. Additional risk factors such as cognitive and functional decline need to be examined for predicting depression risk in later life. Handing and colleagues harness the power of Big Data from the Survey of Health, Ageing, and Retirement in Europe (SHARE Wave 6; $n = 67{,}603$) with machine learning to identify top predictive risk factors for depression in later adulthood.[3] This work presented the first large, multinational study to systematically compare a broad array of socio-relational, health, cognitive, and functional variables as risk factors for depression in middle-aged and older European adults.

The machine learning (ML) approach presented by Handing and colleagues illustrates an effective data-driven framework for testing a large set of potential risk/protective factors for depression. In contrast to a hypothesis-driven approach, ML models can automatically identify patterns and relationships from data without specifying a priori hypotheses. Compared to more conventional statistical methods, ML models make no distributional assumptions and allow researchers to efficiently handle multi-dimensional data and capture the predictive value of all possible combinations of variables in a data set.[4] The random forest analysis used by Handing and colleagues is a supervised ML model well suited to examining interaction and nonlinear effects among a large set of predictors. Top predictive variables can be identified based on feature importance measures, which implicitly capture curvilinear effects and complex interactions. Though powerful for handling many predictors and exploring complex patterns, ML models can be difficult to interpret and don't test statistical significance of predictors. To improve model interpretation, a split-sample methodology was utilized by the authors, which divides data into a subset for training ML models and selecting top predictive variables and another subset for estimating additional parametric models to test statistical significance of the selected predictors.[3] This split-sample strategy is effective in discovering important risk factors and their associations with health outcomes by combining cutting-edge ML models with established statistical methods. Further research is needed on the development and application of interpretable machine learning[5] in predicting depression risk.

Self-reported social isolation and poor health were identified as the strongest risk factors and accounted for over 20% of variability in depression risk.[3] In addition, Handing and colleagues reported that problems with mobility, difficulties in instrumental activities of daily living (in men), and family burden (in women) accounted for approximately an additional 2% of variability in depression risk. These findings point to the need of screening for depression risk in later adulthood during routine health care visits. In particular, middle-aged and older European adults who report being socially isolated are at approximately twice the elevated risk for depression. This suggests that detection of at-risk individuals, a key step in depression prevent and treatment, may be improved by including perceived social isolation measures in the screening process.

Gender difference in the prevalence of depression is well documented.[6] However, it is unclear how different risk/protective factors account for a higher depression rate in women. Handing and colleagues examined a broad range of predictors of depression in middle-aged and older men and women in Europe separately. Despite similar patterns across a number of predictors between women and men, it was found that difficulties in instrumental activities of daily living and self-rated family burden show differential impact in depression risk across sexes.[3] Exploring sex as a potential moderator for linear, nonlinear, and interaction effects of risk/protective factors on depression risk is an interesting topic for future work. Specifically, iterative random forests,[7] a recently developed ML technique, may help to detect stable high-

DOI of original article: http://dx.doi.org/10.1016/j.lanepe.2022.100391

*Corresponding author.
E-mail address: fz53@drexel.edu (F. Zhang).

order interactions among risk/protective factors and sex that are predictive of depression risk.

In summary, Handing and colleagues' work highlights the use of machine learning approach in identifying important risk factors of depression in later adulthood. Future research can build upon this study to examine temporal patterns of risk factors in predicting diagnosis and severity of depression using longitudinal data. A longitudinal study design allows for better control of subject-level heterogeneity, investigation of changes in risk/protective factors and their impact on depression risk, as well as discovery of casual patterns that go beyond cross-sectional associations. In addition, several recent studies investigated machine-learning based brain age estimation from imaging data in depression.[8–10] Integrating clinical data with multimodal neuroimaging data through machine learning may yield new insights about risk factors for depression and merits future research.

Contributors

Both authors contributed equally to the conceptualisation and writing of the manuscript.

Declaration of interests

Both authors declare no competing interests.

References

1 Aichele S, Rabbitt P, Ghisletta P. Illness and intelligence are comparatively strong predictors of individual differences in depressive symptoms following middle age. *Aging Ment Health*. 2019;23(1):122–131.
2 Choi KW, Stein MB, Nishimi KM, et al. An exposure-wide and mendelian randomization approach to identifying modifiable factors for the prevention of depression. *Am J Psychiatry*. 2020;177(10):944–954.
3 Handing EP, Strobl C, Jiao Y, et al. Predictors of depression among middle-aged and older men and women in Europe: A machine learning approach. Lancet Reg Health Eur 2022. https://doi.org/10.1016/j.lanepe.2022.100391.
4 Landrigan J, Zhang F, Mirman D. A data-driven approach to post-stroke aphasia classification and lesion-based prediction. *Brain*. 2021;144(5):1372–1383.
5 Murdoch WJ, Singh C, Kumbier K, et al. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci*. 2019;116(44):22071–22080.
6 Albert PR. Why is depression more prevalent in women? *J Psychiatry Neurosci*. 2015;40(4):219–221.
7 Basu S, Kumbier K, Brown JB, et al. Iterative random forests to discover predictive and stable high-order interactions. *Proc Natl Acad Sci*. 2018;115(8):1943–1948.
8 Dunlop K, Victoria LW, Downar J, et al. Accelerated brain aging predicts impulsivity and symptom severity in depression. *Neuropsychopharmacology*. 2021;46(5):911–919.
9 Han LKM, Dinga R, Hahn T, et al. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Mol Psychiatry*. 2021;26:5124–5139.
10 Niu X, Taylor A, Shinohara RT, et al. Multidimensional brain-age prediction reveals altered brain developmental trajectory in psychiatric disorders. *Cereb Cortex*. 2022.