Machine Learning-Based for Depressive Syndrome

The current polythetic and operational criteria for major depression inevitably contribute to the heterogeneity of depressive syndromes. The heterogeneity of depressive syndrome has been criticized using the concept of language game in Wittgensteinian philosophy. Moreover, "a symptom- or endophenotype-based approach, rather than a diagnosis-based approach, has been proposed" as the "next-generation treatment for mental disorders" by Thomas Insel. Understanding the heterogeneity renders promise for personalized medicine to treat cases of depressive syndrome, in terms of both defining symptom clusters and selecting antidepressants. Machine learning algorithms have emerged as a tool for personalized medicine by handling clinical big data that can be used as predictors for subtype classification and treatment outcome prediction. The large clinical cohort data from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D), Combining Medications to Enhance Depression Outcome (CO-MED), and the German Research Network on Depression (GRND) have recently began to be acknowledged as useful sources for machine learning-based depression research with regard to cost effectiveness and generalizability. In addition, noninvasive biological tools such as functional and resting state magnetic resonance imaging techniques are widely combined with machine learning methods to detect intrinsic endophenotypes of depression.

Keywords: depressive syndrome ; machine learning ; personalized medicine

1. Introduction

Depression is one of the most burdensome disorders worldwide, with a lifetime prevalence of approximately 20% of the global population [1]. Depression remission after the first antidepressant trial is only 30% [2][3]. This low remission rate is partly because diagnosing depression does not guarantee heterogeneous symptom subtypes [4]. Inevitably, the concept that depression is characterized by symptomatic heterogeneity, such as atypical [5], melancholic [6], and anxious [7] subtypes, has gained considerable attention. In addition, it has been reported that the heterogeneity of depressive syndrome can theoretically result from the polythetic and operational criteria of major depression [8][9][10][11][12]. According to the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5) [13], a confirmed diagnosis of major depressive disorder requires both the presence of five or more symptoms among, nine symptoms, including depressed mood, diminished interest or pleasure, weight loss or gain, insomnia or hypersomnia, psychomotor retardation or agitation, fatigue or loss of energy, feelings of worthlessness or excessive guilt, diminished thinking ability or indecisiveness, recurrent thoughts of death or recurrent suicidal ideation, and the presence of either depressed mood or diminished interest or pleasure. Herein, the subset of k draws from n distinguishable objects without replacement and without regard to order that (nCk) can calculate from the theoretical number of different combinations meeting the polythetic and operational criteria of major depressive disorder in DSM-5. Thus, 227 different diagnostic symptom combinations were calculated that can fulfill the DSM-5 diagnostic criteria for major depressive disorder [14][15][16][17]. In terms of psychiatric taxonomy, the heterogeneity of depressive syndrome has been criticized by the concept of a language game in Wittgensteinian philosophy [18]. Wittgenstein suggested the analogy as follows [19]:

Consider for example the proceedings that we call games. I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all?—Don't say: "There must be something common, or they would not be called games"—but look and see whether there is anything common to all.—For if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look!—the concept game is a concept with blurred edges.—"But is a blurred concept a concept at all?"—Is an indistinct photograph a picture of a person at all? Is it even always an advantage to replace an indistinct picture by a sharp one? Isn't the indistinct one often exactly what we need? (Wittgenstein, 2001).

It is also proposed that cases of depressive syndrome are conceptually related by the "family resemblance" rather than the "essence." Thus, it is concluded that the heterogeneity of depressive syndrome is consistent with Wittgensteinian's analogy [18]. Thus, the nomenclature of depressive syndrome can be consistent not with the categorical approach, but the

dimensional approach, in the context of the heterogeneity of major depressive disorder [14]. Furthermore, based on the theoretical construct change from chemical imbalance to dysfunctional circuitry, the symptom-based approach, but not the diagnosis-based approach, has been emphasized by Thomas Insel in his work on the next generation of treatments for mental disorders [20]. Along with the heterogeneity concept, the therapeutic approach also shifts toward selecting antidepressants according to specific symptom clusters [21]. Each cluster of depression symptoms may be thought to react to specific antidepressants, thus potentially improving the current low remission rates. The theorem supporting depression heterogeneity has not generated notable clinical utility in that theory-driven classification of symptom clusters and subsequent antidepressant selection have only produced low accuracies in treatment outcome predictions [22]. However, the clinical utility of the depression heterogeneity concept in diagnostics and therapeutics is increasingly acknowledged with the use of data-driven machine learning approaches.

Machine learning approaches can be more beneficial in the study of depression compared with traditional methods. Factor analysis, for instance, may generate complicated combinations of heterogeneous symptoms within specific dimensions [23]. These analytic approaches also can be vulnerable to experimenter bias in that a researcher has to choose the number of components or clusters in data, as such in k means clustering method [24]. Hierarchical clustering, a type of machine learning method, is an easy-to-implement, deterministic approach in which each of the symptoms is assigned to a single cluster even without predetermining the desired number of clusters.

2. Brain Imaging Techniques and Machine Learning in Depression

Machine learning algorithms are wildly applicable to diverse data of patients for elucidation of the complex nature of depression. In particular, in addition to the aforementioned clinical cohort data, there has been growing attention toward using brain imaging methods to detect endophenotypes of depression that can be clinically significant and feasible for translation to diagnosis [25][26]. Brain MRIs are one of the most widely used techniques that help identify the potential biological markers of depression. Importantly, MRI techniques with machine learning algorithms can produce a classification with brain networks and prediction of treatment response in depression. First, some representative studies adopted graph theory approaches [27][28][29][30] to detect defective functional and structural brain networks of depressed patients. Gong et al. [31] enumerated diverse brain network features, such as alterations in regional and connectivity patterns of different MRI modalities for depression, which include the regional betweenness and degree centrality in structural MRI, region-of-interest-based analysis in functional MRI, and white matter structural connectivity in the diffusion tensor image. Zeng et al. [32] examined the whole-brain functional connectivity at resting state to distinguish the depressed patients from the controls, which yielded 100% sensitivity. The most discriminant functional connectivity was found within or across the affective network, default mode network, and visual cortical areas, which seem to play a critical role in the neuromechanism of depression. Second, other representative studies sought to find alterations in brain network activity at resting state as potential endophenotypes for prediction of therapeutic outcomes. Drysdale et al. [33] suggested that four different patterns of fronto-striatal and limbic functional connectivity be defined as depression biomarkers from functional MRI analyses. The biomarkers were also related to distinctive profiles of clinical symptoms. For instance, biomarker 1 was associated with severe fatigue and anhedonia, and showed best response to repetitive transcranial magnetic stimulation. Redlich et al. [34] examined whether changes in gray matter volume predict response to electroconvulsive therapy. Support vector regression was accompanied with univariate analysis of the Hamilton Depression Rating Scale score, which successfully predicted the response to electroconvulsive therapy and reduction in HDRS. Jiang et al. [35] predicted remission after electroconvulsive therapy using the gray matter of depressed patients, in which six different gray matter networks were suggested as predictors of response to electroconvulsive therapy. Thus, the connectome-based endophenotypes may yield novel opportunities to define the diagnosis of depression and improve therapeutic response.

3. Future Research into Treatment Selection Models

We address the major steps involved in building antidepressant selection models from a clinical database that involves values, for each patient, on variables that represent clinical and demographic characteristics, therapeutics applied to the patient, and observed outcomes from the therapeutics. Understanding the sequential steps is crucial for interpreting and evaluating the utility of findings from the antidepressant selection studies.

The first step was to establish candidate predictor variables. Appropriate candidate predictor variables are those that are acquired prior to the treatment assignment and that credibility could be related to outcome, either generally or differentially between treatments. If a previous study has suggested that a variable can predict an outcome, then it should be involved as a potential predictor variable. However, as the literature on predictors of psychiatric disorders is still relatively scarce, considering other putative variables is recommended.

Variables should be free of significant missingness, and systematic missingness should be examined to ensure the appropriateness of imputation [36]. Variables should also show considerable variability. For instance, it does not make sense to involve sex if 90% of the sample is male. Selecting variables used for prediction is reliant on situations in which predictors exhibit high collinearity. Therefore, it is plausible to test the covariance structure of the putative predictors and take measures to reduce the high collinearity [36]. Other suggestions for identifying putative predictors include addressing outliers, making categorical variables binary, and converting variables for hypothetical reasons or handling highly skewed distributions.

Once putative predictor variables were selected, the next step was to construct the prediction model. This is typically a two-step procedure that includes variable selection and model specifications. Many different variable selection approaches have been suggested for treatment selection, all of which attempt to identify which variables, among the putative predictors, contribute significantly to the prediction outcome. Gillan and Whelan [37] presented an outstanding discussion of data-driven versus theory-driven approaches to model specifications. Typical approaches depend on parametric regression models [38] that select only variables with statistically significant contributions to the outcome. Another approach includes penalties with the goal of limiting the number of selected variables [39]. Others utilize bootstrapping processes that help maximize the generalizability of the models [40][41][42]. Progress in statistical modeling has led to feature selection methods, which are largely based on machine learning algorithms that can compliantly model and identify predictors, even with higher-order interactions [43]. Gillan and Whelan [37] provided an in-depth discussion of the merits of machine learning in the field of psychiatric disorders.

References

1. Vos, T.; Allen, C.; Arora, M.; Barber, R.M.; Bhutta, Z.A.; Brown, A.; Carter, A.; Casey, D.C.; Charlson, F.J.; Chen, A.Z. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016, 388, 1545–1602.

2. Rush, A.J.; Wisniewski, S.R.; Warden, D.; Luther, J.F.; Davis, L.L.; Fava, M.; Nierenberg, A.A.; Trivedi, M.H. Selecting among second-step antidepressant medication monotherapies: Predictive value of clinical, demographic, or first-step treatment features. Arch. Gen. Psychiatry 2008, 65, 870–880.

3. Rush, A.J.; Trivedi, M.H.; Stewart, J.W.; Nierenberg, A.A.; Fava, M.; Kurian, B.T.; Warden, D.; Morris, D.W.; Luther, J.F.; Husain, M.M.; et al. Combining Medications to Enhance Depression Outcomes (CO-MED): Acute and Long-Term Outcomes of a Single-Blind Randomized Study. Am. J. Psychiatry 2011, 168, 689–701.

4. Fried, E. Moving forward: How depression heterogeneity hinders progress in treatment and research. Expert Rev. Neurother. 2017, 17, 423–425.

5. Łojko, D.; Rybakowski, J.K. Atypical depression: Current perspectives. Neuropsychiatr. Dis. Treat. 2017, 13, 2447.

6. Day, C.V.; Williams, L.M. Finding a biosignature for melancholic depression. Expert Rev. Neurother. 2012, 12, 835–847.

7. Ionescu, D.F.; Niciu, M.J.; Mathews, D.C.; Richards, E.M.; Zarate, C.A., Jr. Neurobiology of anxious depression: A review. Depress. Anxiety 2013, 30, 374–385.

8. Park, S.-C.; Kim, J.-M.; Jun, T.-Y.; Lee, M.-S.; Kim, J.-B.; Yim, H.-W.; Park, Y.C. How many different symptom combinations fulfil the diagnostic criteria for major depressive disorder? Results from the CRESCEND study. Nord. J. Psychiatry 2017, 71, 217–222.

9. Park, S.-C.; Kim, D. The centrality of depression and anxiety symptoms in major depressive disorder determined using a network analysis. J. Affect. Disord. 2020, 271, 19–26.

10. Park, S.C.; Jang, E.Y.; Xiang, Y.T.; Kanba, S.; Kato, T.A.; Chong, M.Y.; Lin, S.K.; Yang, S.Y.; Avasthi, A.; Grover, S. Network analysis of the depressive symptom profiles in Asian patients with depressive disorders: Findings from the Research on Asian Psychotropic Prescription Patterns for Antidepressants (REAP-AD). Psychiatry Clin. Neurosci. 2020, 74, 344–353.

11. Kim, Y.-K.; Park, S.-C. An alternative approach to future diagnostic standards for major depressive disorder. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 2020, 105, 110133.

12. Park, S.-C.; Kim, Y.-K. Challenges and Strategies for Current Classifications of Depressive Disorders: Proposal for Future Diagnostic Standards. Major Depress. Disord. Rethink. Underst. Recent Discov. 2021, 1305, 103.

13. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th ed.; American Psychiatric Publishing: Washington, DC, USA, 2014.

14. Østergaard, S.D.; Jensen, S.; Bech, P. The heterogeneity of the depressive syndrome: When numbers get serious. Acta Psychiatr. Scand. 2011, 124, 495–496.

15. Zimmerman, M.; Ellison, W.; Young, D.; Chelminski, I.; Dalrymple, K. How many different ways do patients meet the diagnostic criteria for major depressive disorder? Compr. Psychiatry 2015, 56, 29–34.

16. Park, S.-C.; Kim, Y.-K. Contemporary issues in depressive disorders. Psychiatry Investig. 2019, 16, 633.

17. Park, S.-C.; Kim, Y.-K. Diagnostic issues of depressive disorders from Kraepelinian dualism to the Diagnostic and Statistical Manual of Mental Disorders. Psychiatry Investig. 2019, 16, 636.

18. Rosenman, S.; Nasti, J. Psychiatric diagnoses are not mental processes: Wittgenstein on conceptual confusion. Aust. N. Z. J. Psychiatry 2012, 46, 1046–1052.

19. Wittgenstein, L. Philosophical Investigations, the German Text, with a Revised English Translation; Anscombe, G.E.M., Translator; Blackwell: Malden, MA, USA, 2001.

20. Insel, T.R. Next-generation treatments for mental disorders. Sci. Transl. Med. 2012, 4, ps119–ps155.

21. Uher, R.; Dernovsek, M.Z.; Mors, O.; Hauser, J.; Souery, D.; Zobel, A.; Maier, W.; Henigsberg, N.; Kalember, P.; Rietschel, M. Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. J. Affect. Disord. 2011, 132, 112–120.

22. Arnow, B.A.; Blasey, C.; Williams, L.M.; Palmer, D.M.; Rekshan, W.; Schatzberg, A.F.; Etkin, A.; Kulkarni, J.; Luther, J.F.; Rush, A.J. Depression subtypes in predicting antidepressant response: A report from the iSPOT-D trial. Am. J. Psychiatry 2015, 172, 743–750.

23. Uher, R.; Maier, W.; Hauser, J.; Marušič, A.; Schmael, C.; Mors, O.; Henigsberg, N.; Souery, D.; Placentino, A.; Rietschel, M. Differential efficacy of escitalopram and nortriptyline on dimensional measures of depression. Br. J. Psychiatry 2009, 194, 252–259.

24. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. J. R. Stat. Soc. Ser. C (Appl. Stat.) 1979, 28, 100–108.

25. Patel, M.J.; Andreescu, C.; Price, J.C.; Edelman, K.L.; Reynolds, C.F., III; Aizenstein, H.J. Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. Int. J. Geriatr. Psychiatry 2015, 30, 1056–1067.

26. Haslam, N.; Beck, A.T. Categorization of major depression in an outpatient sample. J. Nerv. Ment. Dis. 1993, 181, 725–731.

27. Yoshida, K.; Shimizu, Y.; Yoshimoto, J.; Takamura, M.; Okada, G.; Okamoto, Y.; Yamawaki, S.; Doya, K. Prediction of clinical depression scores and detection of changes in whole-brain using resting-state functional MRI data with partial least squares regression. PLoS ONE 2017, 12, e0179638.

28. Zhong, X.; Shi, H.; Ming, Q.; Dong, D.; Zhang, X.; Zeng, L.-L.; Yao, S. Whole-brain resting-state functional connectivity identified major depressive disorder: A multivariate pattern analysis in two independent samples. J. Affect. Disord. 2017, 218, 346–352.

29. Wang, X.; Ren, Y.; Zhang, W. Depression disorder classification of fMRI data using sparse low-rank functional brain network and graph-based features. Comput. Math. Methods Med. 2017, 2017, 3609821.

30. Fang, P.; Zeng, L.-L.; Shen, H.; Wang, L.; Li, B.; Liu, L.; Hu, D. Increased cortical-limbic anatomical network connectivity in major depression revealed by diffusion tensor imaging. PLoS ONE 2012, 7, e45972.

31. Gong, Q.; He, Y. Depression, neuroimaging and connectomics: A selective overview. Biol. Psychiatry 2015, 77, 223–235.

32. Zeng, L.-L.; Shen, H.; Liu, L.; Wang, L.; Li, B.; Fang, P.; Zhou, Z.; Li, Y.; Hu, D. Identifying major depression using whole-brain functional connectivity: A multivariate pattern analysis. Brain 2012, 135, 1498–1507.

33. Drysdale, A.T.; Grosenick, L.; Downar, J.; Dunlop, K.; Mansouri, F.; Meng, Y.; Fetcho, R.N.; Zebley, B.; Oathes, D.J.; Etkin, A. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat. Med. 2017, 23, 28–38.

34. Redlich, R.; Opel, N.; Grotegerd, D.; Dohm, K.; Zaremba, D.; Bürger, C.; Münker, S.; Mühlmann, L.; Wahl, P.; Heindel, W. Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. JAMA Psychiatry 2016, 73, 557–564.

35. Jiang, R.; Abbott, C.C.; Jiang, T.; Du, Y.; Espinoza, R.; Narr, K.L.; Wade, B.; Yu, Q.; Song, M.; Lin, D. SMRI biomarkers predict electroconvulsive treatment outcomes: Accuracy with independent data sets. Neuropsychopharmacology 2018, 43, 1078–1087.

36. Jamshidian, M.; Jalal, S. Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. Psychometrika 2010, 75, 649–674.

37. Gillan, C.M.; Whelan, R. What big data can do for treatment in psychiatry. Curr. Opin. Behav. Sci. 2017, 18, 34–42.

38. Fournier, J.C.; DeRubeis, R.J.; Shelton, R.C.; Hollon, S.D.; Amsterdam, J.D.; Gallop, R. Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. J. Consult. Clin. Psychol. 2009, 77, 775.

39. Tibshirani, R. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.) 1996, 58, 267–288.

40. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B (Stat. Methodol.) 2005, 67, 301–320.

41. Austin, P.C.; Tu, J.V. Bootstrap methods for developing predictive models. Am. Stat. 2004, 58, 131–137.

42. Garge, N.R.; Bobashev, G.; Eggleston, B. Random forest methodology for model-based recursive partitioning: The mobForest package for R. BMC Bioinform. 2013, 14, 125.

43. Bleich, J.; Kapelner, A.; George, E.I.; Jensen, S.T. Variable selection for BART: An application to gene regulation. Ann. Appl. Stat. 2014, 8, 1750–1781.