MACHINE LEARNING IN NEUROIMAGING: APPLICATIONS TO BRAIN AGING, AD, SCHIZOPHRENIA, AND BRAIN CANCER

Christos Davatzikos, Ph.D.

Professor of Radiology, Director, Center for Biomedical Image Computing and Analytics, University of Pennsylvania, U.S.A.

Abstract/Summary

Quantitative and computational methods have increasingly provided insights in many neuroscience problems. Amongst them, AI and machine learning have relatively recently emerged as very promising avenues for knowledge discovery, especially in the era of complex, big and diverse data. Herein, applications of machine learning in neuroimaging are discussed, with emphasis on aging and Alzheimer's Disease (AD), schizophrenia, and the most aggressive brain cancer, namely glioblastoma (GBM). In particular, machine learning is shown to produce highly sensitive and specific imaging signatures of brain change during early preclinical stages of AD, as well as to identify neuroanatomically distinct subtypes of schizophrenia. Finally, machine learning is shown to produce imaging signatures that predict patient outcome. These representative results highlight the potential of machine learning in neuroimaging as means to derive sensitive and specific biomarkers, and to reduce complex and diverse data into a small number of dimensions capturing different aspects of the neurobiology of brain diseases.

In the past 30 years we have experienced an exponential growth of various neuroimaging methods, which capture complex aspects of structure, function and connectivity of the human brain, in healthy as well as in diseased states. Quantitative analysis ways have progressed in parallel, responding to the complexity of this data and the richness of the information that can be derived from them. Among these methods, machine learning has emerged as a promising tool for extracting imaging signatures that contribute to personalized precision diagnostics and prognostication [1-6]. Although machine learning has often been viewed as a way to automate tasks that currently require a great deal of human effort (e.g. precise segmentation of anatomical structures or detection of lesions), its greatest potential lies in "seeing" in the data what humans are unable to see, thereby leading to knowledge discovery.

Imaging patterns can be quite complex. For example, no brain region offers sufficient sensitivity and specificity in detecting AD, schizophrenia, and most other brain diseases, despite the fact that numerous studies have associated them with changes in brain volumes, cortical thickness, brain connectivity and function. The main premise of machine learning is that the proper integration of many such "weak predictors" forms strong and highly sensitive and specific imaging signature which can serve as biomarkers of disease and offer personalized prognostications.

This talk discussed two such MR imaging signatures reported in AD [7] and schizophrenia [3], which are identified on an individual basis with promising accuracy. Perhaps most importantly, the former was also found to progressively increase relatively more rapidly in individuals with normal cognition who later progressed to mild cognitive impairment (MCI) [1], thereby potentially offering an early biomarker of AD during stages in which pharmacological and lifestyle interventions might be most effective. In GBM, machine learning derived imaging signatures have been found to improve our predictions of patient outcome [8], thereby offering additional information that can influence patient management, targeted recruitment into clinical trials, as well as more effective evaluation of treatment effects via comparisons to personalized estimated of outcome, rather than to generic population-based medians.

A notorious limitation of machine learning methods has been their often poor generalization and reproducibility in new patients and scans. This weakness is not necessarily fundamental for these methods, but rather emerges from the oftentimes poor application of these approaches to biomedical data. Insufficient training is among the most prominent challenges, as the sheer dimensionality and complexity of various types of neuroimaging data would

normally necessitate training on tens of thousands of scans in order to sample the variability of brain structure and function, as well as to access the diversity of various imaging acquisition protocols and scanner characteristics. Recent work on the formation of international consortia bringing together thousands, or tens of thousands of datasets have offered promise that sufficiently ample and diverse training and validation will be soon possible, which will propel machine learning methods into routine clinical use. Several such consortia were described on studies of brain aging [9, 10], schizophrenia [11, 12], and GBM [13].

As the availability of very large and diverse neuroimaging and clinical datasets increases, additional problems that were previously inaccessible can now be addressed. One such important problem is that of heterogeneity of brain diseases: perhaps seeking an imaging signature of AD or schizophrenia is mundane, since both of these diseases are highly heterogeneous. Recent work has developed advanced semi-supervised machine learning methods, which simultaneously seek to estimate disease subtypes and establish respective imaging signatures [14-16]. Application of these methods to schizophrenia identified two neuroanatomically distinct subtypes/dimensions of schizophrenia, which also showed differences in schizophrenia-related polygenic risk scores. This suggests that diseases that are clinically categorized as unique entities might have quite distinct neuropathological underpinnings, and potentially different response to various treatments. A similar recent study in MCI and AD uncovered 4 dimensions of structural brain change [16], and two progression pathways. Although one of them appeared to be aligned with typical AD-like patterns of atrophy, the second one was more associated to global patterns of brain atrophy potentially related to small vessel ischemic disease and other comorbid pathologies that accelerate the process of brain aging and dementia. Similar work in GBM has identified distinct imaging subtypes of GBM, with differences in patient survival and in molecular characteristics of the tumor [17].

These and many other studies of similar flavor are setting the foundation for more precise definition of neurologic and neuropsychiatric diseases, based on underlying neurobiological signatures, in part derived from imaging. Importantly, such methods are gradually establishing a dimensional view of brain pathologies, with various dimensions informed by neurobiological signatures derived from a variety of biomarkers, including imaging. Eventually, categorizations of a patient into a single and specific disease might become practice of the past, and replaced by placement of a patient in a brain coordinate system spanning the heterogeneity of normal and abnormal brain structure and function. Numerous prior studies can offer contextual information about the clinical implications of a patient being in a particular location on this brain chart (e.g. implications about response to certain treatments). Machine learning methods applied to neuroimaging data gradually and systematically build such dimensions and contextual knowledge.

References

[1] C. Davatzikos, F. Xu, Y. An, Y. Fan, and S. M. Resnick, "Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index", *Brain: a journal of neurology*, vol. 132, no. Pt 8, pp. 2026-35, Aug 2009, doi: 10.1093/brain/awp091.

[2] C. Davatzikos *et al.*, "Classifying spatial patterns of brain activity with machine learning methods: application to lie detection", *NeuroImage*, vol. 28, no. 3, pp. 663-8, Nov 15 2005, doi: 10.1016/j.neuroimage.2005.08.009.

[3] C. Davatzikos *et al.*, "Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities" (in English), *Archives of general psychiatry*, vol. 62, no. 11, pp. 1218-1227, Nov 2005. [Online]. Available: <Go to ISI>://WOS:000233050900005.

[4] S. Kloppel *et al.*, "Automatic classification of MR scans in Alzheimer's disease", *Brain: a journal of neurology*, vol. 131, no. Pt 3, pp. 681-689, March 2008. [Online]. Available: http://brain.oxfordjournals.org/cgi/reprint/awm319v1.

[5] N. Koutsouleris *et al.*, "Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or With Recent-Onset Depression: A Multimodal, Multisite Machine Learning Analysis", *Jama Psychiat*, vol. 75, no. 11, pp. 1156-1172, Nov 1 2018, doi: 10.1001/jamapsychiatry.2018.2165.

[6] N. Koutsouleris *et al.*, "Use of neuroanatomical pattern classification to identify subjects in at-risk mental States of psychosis and predict disease transition", (in eng), *Archives of general psychiatry*, vol. 66, no. 7, pp. 700-712, Jul 2009. [Online]. Available: http://archpsyc.ama-assn.org/cgi/reprint/66/7/700. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19581561.

[7] X. Da *et al.*, "Integration and relative value of biomarkers for prediction of MCI to AD progression: spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers", *NeuroImage. Clinical*, vol. 4, pp. 164-73, 2014, doi: 10.1016/j.nicl.2013.11.010.

[8] L. Macyszyn *et al.*, "Imaging patterns predict patient survival and molecular subtype in glio-

blastoma via machine learning techniques", *Neuro-oncology*, vol. 18, no. 3, pp. 417-25, Mar 2016, doi: 10.1093/neuonc/nov127.

[9] M. Habes *et al.*, "The Brain Chart of Aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans", *Alzheimer's & dementia : the journal of the Alzheimer's Association*, vol. 17, no. 1, pp. 89-102, Jan 2021, doi: 10.1002/alz.12178.

[10] R. Pomponio *et al.*, "Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan", *NeuroImage*, vol. 208, p. 116450, 2020.

[11] G. B. Chand *et al.*, "Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning", *Brain: a journal of neurology*, vol. 143, no. 3, pp. 1027-1038, Mar 1 2020, doi: 10.1093/brain/awaa025.

[12] M. Rozycki *et al.*, "Multisite Machine Learning Analysis Provides a Robust Structural Imaging Signature of Schizophrenia Detectable Across Diverse Patient Populations and Within Individuals", *Schizophrenia bulletin*, vol. 44, 11/24 2017, doi: 10.1093/schbul/sbx137.

[13] C. Davatzikos *et al.*, "AI-based prognostic imaging biomarkers for precision neuro-oncology: the ReSPOND consortium", *Neuro-oncology*, vol. 22, no. 6, pp. 886-888, Jun 9 2020, doi: 10.1093/neuonc/noaa045.

[14] A. Dong *et al.*, "Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers," (in eng), *Brain: a journal of neurology*, vol. 140, no. 3, pp. 735-747, Mar 01 2017, doi: 10.1093/brain/aww319.

[15] E. Varol, A. Sotiras, C. Davatzikos, and I. Alzheimer's Disease Neuroimaging ,"HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework" (in Eng), *NeuroImage*, vol. 145, no. Pt B, pp. 346-364, Jan 15 2017, doi: 10.1016/j.neuroimage.2016.02.041.

[16] Zhijian Yang, Junhao Wen, and C. Davatzikos, "Smile-GANs: Semi-supervised clustering via GANs for dissecting brain disease heterogeneity from medical images", *Arxiv*, vol. arxiv.org/abs/2006.15255, 2020.

[17] S. Rathore *et al.*, "Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1", *Nature Scientific Reports*, vol. 8, no. 1, p. 5087, 2018/03/23 2018, doi: 10.1038/s41598-018-22739-2.