Supervised machine learning in psychiatry

Contributions

The contributions of this PhD thesis are:

- the initial development of SML algorithms for the prediction for the 3-year prediction of conversion from Mild Cognitive Impairment to Alzheimer's Disease (Chapter 2-4), and the prediction of 2-year Obsessive-Compulsive Disorder remission (Chapter 5)
- the development of the abovementioned algorithms by using only limited cost and clinically accessible predictive information (Chapter 2-5)
- the introduction of different validation and testing protocols that, even in early development phases, allow providing more accurate estimates of the expected algorithm performance when applied in clinical practice and in multiple clinical contexts, with the aim to steering the development of SML algorithm towards a clinical applicability since the initial phases (Chapter 3-5)
- the extensive clinical validation of a SML algorithm that can automatically perform sleep staging of polysomnography (Chapter 6)

Impact Paragraph

In recent years, there has been an exponential growth of scientific literature regarding the application of SML in Psychiatry. This interest has been motivated by the promise of using SML to develop new clinical tools that could help to perform personalized predictions and recommendations, ultimately improving the results achievable in the psychiatric clinical practice. Starting from the evidence of a substantial lack of such tools in Psychiatry, the studies presented in this dissertation aimed to contribute to further directing the application of SML towards the original promise. In particular, they demonstrate that it is possible to develop SML algorithms that reach clinically relevant performances even by employing only input variables that are or may be easily accessible in the clinical routine, and avoiding those that are still too expensive, invasive, or hard to introduce into clinical psychiatric practice. This selection of the input information is crucial to prevent SML algorithms to

remain just promising proofs-of-concept with limited opportunity to become applicable in practice.

Moreover, the studies also contributed to highlighting the importance of providing estimates of the generalized performance of an SML algorithm even at early development phases. This implies an investigation of what is the expected performance of an SML algorithm when applied in totally new cases, as well as in new clinical settings. This is a necessity because no clinical application of a medical device can be made before a thorough investigation of its safety and efficacy. Doing it systematically at every step of the development process allows to early identify any generalizability issue and to promptly act to solve it along the entire development process. The studies in this thesis introduced peculiar performance testing strategies, specifically designed based on the level at which the development of the algorithm was, the nature of the task under study, and the data available. These strategies may also be used in other studies with similar characteristics or inspire innovative testing protocols.

Overall, the results of the studies included in this doctoral dissertation contribute to demonstrating that the use of SML algorithms in psychiatric clinical practice is not just a promise, even though the process to reach a practical application may require several redesigns of the algorithms and significant evidence in support of their efficacy. Psychiatry may substantially benefit from a shift towards a Personalized Medicine approach to improve the prevention and treatment of mental disorders, which still have significant margins of improvement. Thus, the potential progress achieved in the clinical practice may be worth all the efforts required to complete the development and clinical validation of an SML algorithm. The advantage of using SML is that it does not require an explicit understanding of the phenomena under investigation, but rather the availability of enough suitable examples to be used to train the algorithm. The studies presented in this thesis show how the use of SML may enable to perform psychiatric clinical tasks that were only in part possible previously, e.g., an early prediction of conversion to Alzheimer's Disease in high-risk individuals or of remission in subjects suffering from Obsessive-Compulsive Disorder. The study in Chapter 6 also demonstrates how SML may allow to speed up some clinical procedures

in clinical practice, reducing the costs and the associated clinicians' burden.

This doctoral project has been conceived as inherently multidisciplinary. A joint work among multiple parties and professional figures is necessary to develop SML algorithms that aim to become clinically used tools. Machine Learning requires theoretical and technical skills beyond the average expertise of the typical research scientist in Psychiatry. At the same time, machine learning experts need to work closely and continuously with domain experts from both the research and the clinical side to receive directions regarding which tasks may be relevant to address with SML, which available scientific knowledge can be used to better design and improve the algorithms, and which constrains needs to be satisfied to make them effectively applicable in practice. Besides, further experts need to be involved to effectively transform a SML algorithm into a usable clinical tool, e.g., software engineers, user-experience designers, and regulatory specialists. These interdisciplinary collaborations may foster additional exchanges beyond the sole activities regarding SML, ultimately promoting the beginning of new projects and innovative ideas in all the involved disciplines.

Finally, all studies presented in this doctoral dissertation have been performed with the collaboration of different research groups and institutions, and this doctoral work contributed to further strengthening existing partnerships as well as creating new ones[14]. Part of these collaborations revolved around the sharing of privately-owned datasets that have been used for the development of SML algorithms for the first time. Several psychiatric datasets suitable to be employed for this purpose may exist, but they may not be publicly available, and they may have never been used in SML projects before. The studies presented in this dissertation may also contribute to foster a new use of already available datasets that ultimately will ease the beginning of new SML projects in Psychiatry and make a larger number of institutions and researchers in the psychiatric field approach SML for the first time.

14 The main research groups and institutions involved in this doctoral project were: the School for Mental Health and Neuroscience (MHeNs) and the Institute of Data Science (IDS) of Maastricht University, Villa San Benedetto Menni Hospital (Albese con Cassano, CO, Italy), Humanitas University (Rozzano, MI, Italy), Mount Sinai Medical Center and Miami University (Miami, FL, USA), the Netherlands Obsessive Compulsive Disorder Association (NL), and Medibio Limited (Savage, MN, USA).