Using Machine Learning on mHealth-based Data Sources

Rüdiger Pryss[1(✉)], Marc Schickler[2], Johannes Schobel[3], Winfried Schlee[4], Myra Spiliopoulou[5], Thomas Probst[6], and Felix Beierle[1]

[1] Institute of Clinical Epidemiology and Biometry, University of Würzburg, 97080 Würzburg, Germany
{ruediger.pryss,felix.beierle}@uni-wuerzburg.de
[2] Institute of Databases and Information Systems, Ulm University, 89081 Ulm, Germany
marc.schickler@uni-ulm.de
[3] Institute DigiHealth, Neu-Ulm University of Applied Sciences, 89231 Neu-Ulm, Germany
johannes.schobel@hnu.de
[4] Department of Psychiatry and Psychotherapy, University of Regensburg, 93053 Regensburg, Germany
winfried.schlee@gmail.com
[5] Knowledge Management and Discovery Lab, Otto-von-Guericke University, 39106 Magdeburg, Germany
myra@ovgu.de
[6] Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems an der Donau, Austria
thomas.probst@donau-uni.ac.at

Abstract. The application of machine learning algorithms has become important for the medical domain. However, the concrete application of these type of algorithms strongly depends on how a corresponding data source was created. Most importantly, domain knowledge must be linked with data science knowledge. Data collected using smartphones or smart mobile devices (e.g., smart watches) is commonly referred to as mHealth data. The possibilities and strategies for collecting data in this area now appear to be as diverse as the machine learning algorithms that have emerged. This tutorial will therefore discuss how mHealth data is structured and which aspects need to be taken into account when evaluating it with machine learning algorithms, using concrete examples.

Keywords: mHealth · Machine learning · Data collection strategies

1 Background Information and Tutorial Content

The use of smartphones - and, by extension, smart mobile devices - for clinical trials is no longer a niche phenomenon. Above all, it has been recognized that smartphones can measure data in-situ in order to achieve a high transferability of

M. Michalowski et al. (Eds.): AIME 2022, LNAI 13263, pp. 443–445, 2022.
https://doi.org/10.1007/978-3-031-09342-5

the results to reality; also known as high ecological validity of the data. Of note, only a few standards have been established in this area so far, so that this domain is still characterized by many utilized data collection strategies and also new developments [1]. Another circumstance complicates the situation, the utilized or developed concepts that are currently used come from different scientific domains and therefore often make little use of each other's achievements. To get a better impression, Fig. 1 shows the predominant concepts and which scientific domain they originate from.

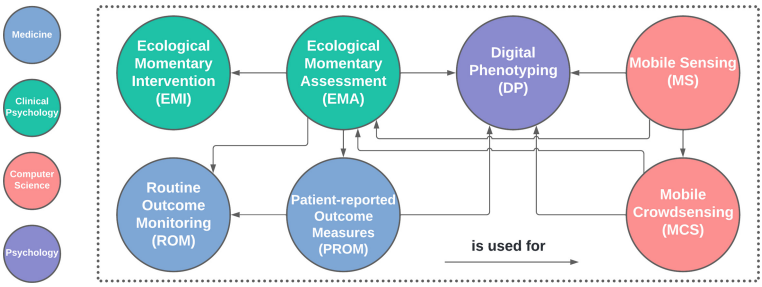

Fig. 1. mHealth Research and Data Collection Concepts

As can be seen in Fig. 1, currently seven major flavors can be distinguished to collect mHealth data. Some of these are combined, and some have additional sub-flavors. The implementation of these strategies for clinical studies mainly leads to two consequences. First, comparing study data is often very difficult, as it is rarely specified exactly which paradigm was used. Second, the strategies lead to the fact that if a study design is not strictly clinically supervised, which is not wanted in the mHealth context most of the time (i.e., ecological validity can only be achieved when data is gathered in the wild), then participants fill in data in very different frequencies, intervals and quantities [2]. These different distributions ultimately lead to the fact that evaluations by means of machine learning should take very many bias types into account in order to achieve meaningful results. For the reasons mentioned above, it is then very often argued that there is simply too little data available to be able to make meaningful statements. From our project experience, this view is not proper in many respects.

To summarize, many aspects play a role that currently need to be considered when machine learning is applied to mHealth data. Moreover, there are many types of software development flavors (e.g., native vs. cross-development) that also play a role in the sketched data collection settings. Considering all of the mentioned aspects, we want to discuss in this tutorial along existing evaluations on mHealth data, which possibilities of addressing them exist and how a suitable interpretation of the data using machine learning can look like.

References

1. Kraft, R., et al.: Combining mobile crowdsensing and ecological momentary assessments in the healthcare domain. Front. Neurosci. **14**, 164 (2020). https://doi.org/10.3389/fnins.2020.00164
2. Schleicher, M., et al.: Understanding adherence to the recording of ecological momentary assessments in the example of tinnitus monitoring. Sci. Rep. **10**, 22459 (2020). https://doi.org/10.1038/s41598-020-79527-0