Using machine learning in aging neuroscience

By

Christian Johannes Gölz

A thesis presented for the degree of
Doctor rerum naturalium
(Dr. rer. nat)

Paderborn University
Faculty of natural sciences
2022

Contents

Acknowledgement

Abstract

Aim: Apply data science methods to questions in aging Neuroscience
Methods: Supervised and unsupervised methods in different settings
Results: Novel Data Driven insights
Coclusion: ML rocks!

Figures

Tables

List of Abbreviations

Publications and other scientific contributions

Chapter 1

General Introduction

1.1 Motivation

Driven by ever-increasing amounts of data and advancing computer infrastructure, the field of artificial intelligence (AI) is becoming increasingly influential in socially important areas. These include public transportation, e.g., autonomous vehicles [1], the medical sector, e.g, diagnostic imaging [2] or social interaction, e.g., conversational AI or so-called chatbots [3]. Likewise, progress in science is more and more characterized by the application of methods from AI such as algorithms from Machine Learning, which makes it possible to analyze large and complex amounts of data systematically [4]. This has led to proclamations of an ”AI revolution in science” [5] or promoting science has entered the fourth paradigm characterized by *data-intensive computing* [6].

Thus, AI as a key technology becomes a hope for solving societal challenges. One of the greatest challenges in industrialized countries is the demographic shift toward an older population. This poses enormous challenges for society as a whole raising issues for the healthcare system, infrastructure, family policy, and the occupational sector. To avoid overloading these social structures, one of the main goals is the promotion of healthy, independent aging and reduction of the impact on the healthcare system by use of smart technologies for assistance in daily life, rehabilitation or care.

In general aging is an ongoing process that can be detected at multiple interacting biological systems operating on several spatial and temporal scales contributing to the complexity of the phenomenon [7]. The most prominent consequences of this are declines in cognitive and sensorimotor abilities challenging the daily life of older adults. Understanding and describing the neurophysiological underpinnings is crucial to distinguish age-related diseases from healthy aging at an early stage, to develop and plan targeted interventions, or to develop smart technologies adapted to an aging population. Besides the early automatic detection of pathological changes, tools from AI and especially Machine Learning can be used to describe individual trajectories of aging helping to identify biomarkers that could be used to verify treatments as well as targeted interventions. Other applications include systems to assist older adults in daily living or care assistive technology, such as robotic tools for neurorehabilitation. However, such applications and smart technologies are still under development.

Acceptance and targeted use of AI based technology requires piloting of application areas in science and practice. In particular, it is unclear to what extent age-related changes can be modelled by algorithms from the field of Machine Learning and, conversely, how smart technologies and age-related changes may interfere. Therefore, the aim of this thesis is to identify and apply Machine Learning techniques to address age-related changes such as the neurophysiological underpinnings and influencing factors of sensorimotor and cognitive decline. The intention is to test to what extent hypotheses can be confirmed, new hypotheses can be formed and derivations for the development of targeted interventions and AI-based smart technology can be made.

1.2 Outline of the thesis

This thesis is separated in six main chapters. In this chapter (chapter 1) the framework of this thesis is described. This includes a theoretical and methodological introduction to machine learning and aging with a focus on neuroscience aspects (section 1.3). General terminology as well as a literature-based overview of the use of machine learning methods in neuroscience and especially in the neuroscientific research on aging will form the basis for the deduction of research aim and scope of this thesis in the following section 1.4. Furthermore, this shall serve as a background for considerations and the description of the methodological approach of this work (section 1.5). In the subsequent chapters 2 to 5 the following published subprojects underlying this thesis will be presented:

- chapter 2: Goelz, C. *et al.* Classification of visuomotor tasks based on electroencephalographic data depends on age-related differences in brain activity patterns. *Neural Networks* **142** (May 2021)
- chapter 3: Goelz, C. *et al.* Using machine learning to characterize electrophysiological correlates of selective attention across the lifespan. *unpublished* (2022)
- chapter 4: Goelz, C. *et al.* Electrophysiological signatures of dedifferentiation differ between fit and less fit older adults. *Cognitive Neurodynamics* **15,** 1–13 (Oct. 2021)
- chapter 5: Gaidai, R. *et al.* Classification characteristics of fine motor experts based on electroencephalographic and force tracking data. *Brain Research* **1792,** 148001 (July 2022)

The thesis concludes with an overreaching discussion highlighting consequences and future research topics (chapter 6).

1.3 Theoretical Background and terminology

The following chapters include an introduction to Machine Learning along with basic formalism and terminology. Next an overview of how Machine Learning is used in the field of Neuroscience will be given by highlighting relevant literature. Subsequently, the changes in the aging brain that are relevant for this thesis will be presented as well as the application of Machine Learning methods used in this field. The goal is to highlight application areas and methodologies that are relevant for the following research.

1.3.1 Machine learning

Machine learning emerged in the 1950s as a subbranch of artificial intelligence to enable computers to learn without being explicitly programmed [12]. It is defined by algorithms that automatically extract patterns and trends or *learn* from data [13]. The notion of *learning* therein describes the process of acquiring the ability to generalize these trends and patterns to unknown data not used during the process. The goal of Machine Learning is therefore to extract generalizable patterns based on examples or so-called training data that allow new data to be classified or predictions to be made. The data may contain few or multiple properties, so-called features, and may be multidimensional with variable sources for example sensor recordings or pixel values.
Besides solving computational problems algorithms from Machine Learning offer additional value for scientific inquiry including the possibility for the automatic analysis of complex multidimensional data [4, 14]. In this approach, rather than assuming that data are generated by an underlying stochastic model as in classical statistical modelling the data mechanisms are treated

as unknown here which may overcome inaccuracies in the analysis [14]. In this notion a model can be seen as a mathematical formulation of nature

Furthermore, by extracting generalizable principles from the complex interaction of features it offers additional values to traditional hypothesis-driven approaches [15, 16].

Machine learning can be subdivided into three categories, supervised, unsupervised and reinforcement learning. In supervised machine learning the goal is to learn a model representing the relationship between data and associated information or description, a so-called label or target. This model can than be used to predict the label of new data that not have been used during model creation.

If the labels are categorical, this process is called classification; for continuous labels, the term is regression. The goal of unsupervised machine learning is to find hidden structure in data without taking into account associated labels. This could be grouping similar data points, i.e. clustering, or uncovering a meaningful low dimensional representation of the data, i.e. dimensionality reduction. Rather than a specific label in reinforcement learning the goal is to learn optimal actions to solve a certain problem by maximizing the reward linked to that action.

1.3.1.1 Applications on Neuroscience

Aiming at building an artificial general intelligence machine learning algorithms and technology were inspired by the working principle of the brain leading to advances of neural networks [17] and neuromorphic computing [18]. As already introduced machine learning on the other side offers great tools for a wide variety of problems addressed areas of neuroscience.

- Clinical: identify disease, develop biomarkers, characterize patients, epilepsy detection/prediction - Basic: Understand working principle of processing, e.g. visual system, working memory - Cognitive: Identify brain states and study brain behavior interaction

Summary: - Solving engineering problems as well as understanding brain processing - Investigate high dimensional representations with classification/regression/model selection - Uncover underlying processes with dimensionality reduction

1.3.2 Age related reorganization of the brain

Age related reorganization processes are detectable at the whole body. This is underpinned by multiple interacting biological systems operating on several spatial and temporal scales contributing to the complexity of the phenomenon [7]. At the behavioral level these processes are noticeable in changes in cognitive, motor and sensory functioning [QUELLE]. Aging is one of the biggest risk factors for neurodegenerative diseases such as dementia, including Alzheimer's disease, as well as Parkinson's disease making the brain as one of the target systems to study. Patterns of reorganization of the brain are highly individual as they are subject to genetic and environmental influences [QUELLEN]. At the same time, however, overarching, generalizable patterns can be detected [QUELLE]
On a structural level aging has been associated with a reduction in gray matter with an onset early in life

1.3.2.1 Machine Learning usages in aging neuroscience

1.4 Aims and scope

The complex interplay of the aforementioned factors leading to the dynamics of brain network changes over the lifetime perspective is not fully understood. However, this is a prerequisite to differentiate healthy from pathological changes and to develop and verify treatments as well as targeted interventions. This requires uncomplicated, easy-to-use, and cost-effective methods and novel analyses to quantify changes in brain organization. Several methods are available to study the brains' structure and function including functional and structural Magnetic Resonance Imaging (s/fMRI) as well as Magnetoencephalography (MEG) and Electroencephalography (EEG). In addition to cost-intensive imaging techniques such as fMRI, EEG, which meets these criteria, is particularly suitable. Moreover, EEG measures neuronal activity directly with high temporal resolution, which would allow us to gain new insights into the reorganization of brain networks over the lifespan in health and disease. However, it is unclear how the changes described above are reflected in electrophysiological markers. Furthermore, EEG signals are temporally and spatially highly dimensional and have a low signal-to-noise ratio, which makes the detection and visualization of brain networks and their dynamics difficult and requires advanced signal analysis methods. Advanced methods such as methods from the field of artificial intelligence with machine learning, as a subbranch, are of special interest in this context. Methods from supervised and unsupervised machine learning are possible candidates. In unsupervised machine learning, the goal is to find structure in the data. This includes methods for dimensionality reduction and clustering. Dimensionality reduction for example allows us to describe the structure of high dimensional data in fewer properties [20]. Two common methods in the analysis of neurophysiological data are the principal component analysis (PCA) and independent component analysis (ICA). These allow the detection of spatial patterns in the data that represent the underlying network characteristics of neurophysiological data [8]. In addition, with dynamic mode decomposition (DMD), Brunton, Johnson, et al. [21] apply for the first time a method to electrophysiological data that allows us to map both the spatial and temporal structure of the network structure of neurophysiological data. In supervised machine learning, models are created that can predict a certain outcome based on input data. This method is used to detect neuronal representations of the environment or certain behaviors as well as group memberships and to identify relevant markers [22]. In the context of lifespan changes, complex brain network behavior based on EEG data could be extracted and visualized using dimensionality reduction. Supervised machine learning methods could be used to detect representations of the environment and behavior and to draw conclusions about the differentiation of brain networks. Automatic detection of group membership could further provide new predictors of nervous system states.

1.5 General methodology

Chapter 2

Paper 1

Chapter 3

Paper 2

Chapter 4

Paper 3

Chapter 5

Paper 4

Chapter 6

General discussion

Bibliography

1. Leonard, J. J., Mindell, D. A. & Stayton, E. L. Autonomous vehicles, mobility, and employment policy: the roads ahead. *Massachusetts Inst. Technol., Cambridge, MA, Rep. RB02-2020* (2020).
2. Liu, Y., Jain, A., Eng, C., Way, D. H., Lee, K., Bui, P., Kanada, K., de Oliveira Marinho, G., Gallegos, J., Gabriele, S., Gupta, V., Singh, N., Natarajan, V., Hofmann-Wellenhof, R., Corrado, G. S., Peng, L. H., Webster, D. R., Ai, D., Huang, S. J., Liu, Y., Dunn, R. C. & Coz, D. A deep learning system for differential diagnosis of skin diseases. *Nature Medicine* **26,** 900–908. ISSN: 1546-170X. https://doi.org/10.1038/s41591-020-0842-3 (June 2020).
3. Adamopoulou, E. & Moussiades, L. *An Overview of Chatbot Technology* in *Artificial Intelligence Applications and Innovations* (eds Maglogiannis, I., Iliadis, L. & Pimenidis, E.) (Springer International Publishing, Cham, 2020), 373–383. ISBN: 978-3-030-49186-4. https://doi.org/10.1007/978-3-030-49186-4_31.
4. Brunton, B. W. & Beyeler, M. Data-driven models in human neuroscience and neuroengineering. *Current Opinion in Neurobiology* **58.** Computational Neuroscience, 21–29. ISSN: 0959-4388. https://www.sciencedirect.com/science/article/pii/S0959438818302502 (2019).
5. Appenzeller, T. *The AI revolution in science: How deep learning is helping scientists cope with a data deluge* 2017. https://www.science.org/content/article/ai-revolution-science.
6. Hey, T., Tansley, S., Tolle, K. & Gray, J. *The Fourth Paradigm: Data-Intensive Scientific Discovery* ISBN: 978-0-9825442-0-4. https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/ (Microsoft Research, Oct. 2009).
7. Mooney, K. M., Morgan, A. E. & Mc Auley, M. T. Aging and computational systems biology. *WIREs Systems Biology and Medicine* **8,** 123–139. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wsbm.1328. https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wsbm.1328 (2016).
8. Goelz, C., Mora, K., Rudisch, J., Gaidai, R., Reuter, E.-M., Godde, B., Reinsberger, C., Voelcker-Rehage, C. & Vieluf, S. Classification of visuomotor tasks based on electroencephalographic data depends on age-related differences in brain activity patterns. *Neural Networks* **142** (May 2021).
9. Goelz, C., Reuter, E.-M., Fröhlich, S., Rudisch, J., Vieluf, S. & Voelcker-Rehage, C. Using machine learning to characterize electrophysiological correlates of selective attention across the lifespan. *unpublished* (2022).
10. Goelz, C., Mora, K., Stroehlein, J., Haase, F., Dellnitz, M., Reinsberger, C. & Vieluf, S. Electrophysiological signatures of dedifferentiation differ between fit and less fit older adults. *Cognitive Neurodynamics* **15,** 1–13 (Oct. 2021).

11. Gaidai, R., Goelz, C., Mora, K., Rudisch, J., Reuter, E.-M., Godde, B., Reinsberger, C., Voelcker-Rehage, C. & Vieluf, S. Classification characteristics of fine motor experts based on electroencephalographic and force tracking data. *Brain Research* **1792,** 148001 (July 2022).

12. Samuel, A. L. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development* **3,** 210–229 (1959).

13. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Second Edition* Second. ISBN: 978-0-387-84857-0. https://link.springer.com/book/10.1007/978-0-387-84858-7 (Springer-Verlag, New York, 2009).

14. Breiman, L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science* **16,** 199–231. https://doi.org/10.1214/ss/1009213726 (2001).

15. Vu, M.-A. T., Adalı, T., Ba, D., Buzsáki, G., Carlson, D., Heller, K., Liston, C., Rudin, C., Sohal, V. S., Widge, A. S., Mayberg, H. S., Sapiro, G. & Dzirasa, K. A Shared Vision for Machine Learning in Neuroscience. *Journal of Neuroscience* **38,** 1601–1607. ISSN: 0270-6474. eprint: https://www.jneurosci.org/content/38/7/1601.full.pdf. https://www.jneurosci.org/content/38/7/1601 (2018).

16. Bzdok, D. & Yeo, B. T. Inference in the age of big data: Future perspectives on neuroscience. *NeuroImage* **155,** 549–564. ISSN: 1053-8119. https://www.sciencedirect.com/science/article/pii/S1053811917303816 (2017).

17. Macpherson, T., Churchland, A., Sejnowski, T., DiCarlo, J., Kamitani, Y., Takahashi, H. & Hikida, T. Natural and Artificial Intelligence: A brief introduction to the interplay between AI and neuroscience research. *Neural Networks* **144,** 603–613. ISSN: 0893-6080. https://www.sciencedirect.com/science/article/pii/S0893608021003683 (2021).

18. Choi, C., Kim, H., Kang, J.-H., Song, M.-K., Yeon, H., Chang, C. S., Suh, J. M., Shin, J., Lu, K., Park, B.-I., *et al.* Reconfigurable heterogeneous integration using stackable chips with embedded artificial intelligence. *Nature Electronics* **5,** 386–393. ISSN: 2520-1131. https://doi.org/10.1038/s41928-022-00778-y (2022).