COMMUTE
Quarterly Bulletin
October 2024

Prof. Dr. Martin Hofmann-Apitius, Coordinator and Colleague

Introducing Work Package 3 – the Data Science & AI activities

In COMMUTE, we bring together a very heterogeneous group of scientists to address one of the most challenging aspects of the COVID-19 pandemic: the role of SARS-CoV-2 in triggering processes that ultimately lead to neurodegeneration.

In this bulletin of COMMUTE, we introduce the "Data-Driven Hypothesis-Free Approach" that we organize in Work Package (WP) 3.

The COMMUTE WP structure is intentionally kept simple, and as a consequence, we have bundled all activities of all partners with expertise in Data Science and Artificial Intelligence in WP3. We call this WP "Applied Artificial Intelligence" because the focus is not on developing AI methods but rather on applying existing Data Science and AI methods to find patterns and signals that represent causal relationships and that are mechanistically interpretable. The resulting AI models will allow for the assessment of the personal risk of developing neurodegenerative diseases (NDDs), such as Alzheimer's disease (AD) or Parkinson's disease (PD) after a SARS-CoV-2 infection. Of course, the available and usable data must comprise the relevant information – what has not been measured cannot be tested.

Data Science and AI activities in COMMUTE have two major goals:

1. Develop testable comorbidity hypotheses based on identified interpretable signals and patterns, and
2. Generate individual-level risk models (plural!) for the prototype implementation of a recommender system (clinical decision support).

The COMMUTE data catalog

When we designed COMMUTE, we were fully aware that the comorbidity between COVID-19 and Neurodegenerative Diseases may well be represented in different types of data and that we would need to analyze data representing different modalities and different levels. It was clear from the beginning that we would need not only comprehensive data sets representing diverse modalities, but also the expertise to analyze these different data.

Therefore, already at the level of the grant application, the COMMUTE proposal presented a complete data inventory indicating all different types and modalities of data accessible to and usable by individual consortium members.

This overview of various data and the research questions that can be addressed to these data sets is currently being implemented as a searchable resource: the COMMUTE data catalog. This catalog will comprise a comprehensive, searchable index of multimodal data used in COMMUTE WP3, accompanied by harmonized, rich metadata. The COMMUTE data catalog will also be shared amongst the "sister projects" of COMMUTE (there are eight other projects funded in the same call, and many of them deal with the comorbidity between viruses and various chronic diseases).

Let us now look at the teams and their expertise that make WP3. Since AI approaches are closely linked to the data accessible and usable for the different teams, we will introduce not only names and institutions but also some details on how they contextualize data and what modeling and mining approaches they apply.

Real-World and Population Data from Denmark:
Søren Brunak and his team at the University of Copenhagen

Søren is a leading scientist in bioinformatics and health informatics with a strong focus on translating findings from the population to the individual level. His team's previous work has impressively demonstrated that detailed insights on medical conditions and comorbidities can be extracted from Danish population-wide real-world data; this approach is now being adapted to study the link between COVID-19 and neurodegenerative diseases. Søren's team will use modern Machine Learning approaches to detect patterns that are associated with accelerated neurodegenerative disease progression and development in the Danish population by comparing pre- and post-pandemic incidences and prevalences.

Real World Data with a focus on Environmental Factors from Catalonia:

Alfonso Valencia and his team at the Barcelona Supercomputing Center (BSC)

Alfonso, Director of the Life Sciences Department of BSC, a major research institution in Europe focused on high-performance computing (HPC), is a leading computational biologist and bioinformatician known for his work in systems biology, Artificial Intelligence, and big data analysis in the life sciences. His particular interest in the influence of environmental factors on COVID-NDD comorbidity risk has prompted his team to focus on real-world population data from Catalonia. Jon Sanchez Valle, postdoctoral researcher, and Davide Cirillo, head of the Machine Learning for Biomedical Research Unit at BSC, will model the influence of environmental factors on the risk of severe COVID-19 and – as linked comorbidity – the increased risk for neurodegenerative disease initiation and progression using causal Machine Learning and explainable AI. The data they will use encompass clinical information from electronic health records managed by the Catalan Health Institute and high-resolution air quality forecasts using dedicated models developed at BSC. All these resources will be complemented by integrating information from reputable external public databases of molecular information, including pathways and drug targets, to provide more comprehensive models.

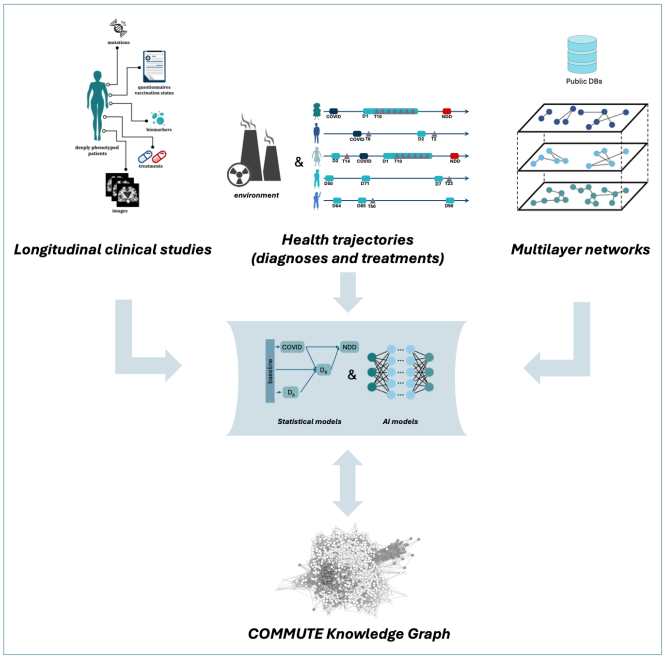

Illustration of the AI approach taken by partner BSC
Longitudinal clinical studies and real-world data (EHRs with Health Trajectories) are combined with a priori knowledge (Multilayer Networks) for integrative modeling.
Patterns (and signals) resulting from the modeling approach contribute to the COMMUTE knowledge graph.

Global Population Health Data and the Role of Contextual Socio-Psychological Factors from Oxford:

Sarah Bauermeister and her team at the University of Oxford

Professor Sarah Bauermeister is co-director of the Dementias Platform UK (DPUK)-Korea Brain Research Institute (KBRI) Joint Research Centre, principal investigator of the Early Adversity and Brain Health Programme 'Blossom' at the University of Oxford, and part of the COVID Global Mental Health Consortium (CGMHC) coordinated by Prof. Jordan Smoller at Harvard Medical School. Together with her team at the University of Oxford, Prof. Sarah Bauermeister brings to this project her expertise in adversity and brain health and her expertise in the application of longitudinal analyses, in particular Structural Equation Modeling (SEM), to large-scale longitudinal multimodal data. In this project, Prof. Sarah Bauermeister is investigating how contextual socio-psychological factors (i.e., early adversity) inter-influence and interact with mental health before, during, and after the COVID-19 pandemic to contribute to the overall risk of COVID-19 neurodegeneration. The methodology used by Professor Bauermeister has a proven track record in epidemiology and public health. It will enable us to study and understand the underlying dynamics contributing to the risk of developing neurodegenerative disease after a SARS-CoV-2 infection.

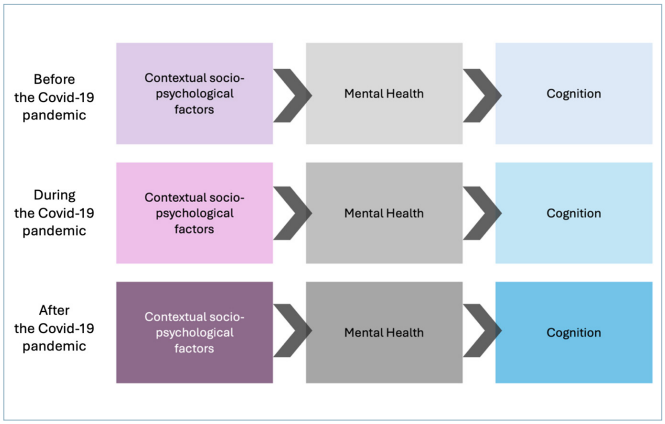

Illustration of the approach developed by the Oxford Team around Prof. Sarah Bauermeister
Socio-psychological factors impacting mental health will be studied in relation to the period before, during, and after the pandemic. The systematic comparison of large data sets allows for the identification of socio-psychological factors that contribute to personalized risk for people exposed to the virus under sociological and psychological conditions.

Utilizing Big Data in Health:

Holger Fröhlich and his team at the Fraunhofer Institute for Algorithms and Scientific Computing (SCAI)

Holger Fröhlich and his team at the Fraunhofer Institute for Algorithms and Scientific Computing (SCAI) are using advanced AI on big data from commercial vendors (IBM, TriNetX) and other sources (UK BioBank) to model risk (in the sense of "time-to-event" models) at the level of individuals, depending on medical history as encoded in electronic health records. They determine and compare baseline risks of developing NDDs in the population before, during, and after the pandemic. By matching comparable subjects with and without a reported COVID-19 infection and using recent causal Machine Learning techniques, the team estimates the effect of a COVID-19 infection on NDD risk. Holger Fröhlich and his team have been working in that space before and will specifically look for explainable patterns that can be used for prediction on one side and for the generation of testable hypotheses on the other side. They use methods like SHAP (SHapley Additive exPlanations) and related methodologies to make their models explainable and testable at the same time.

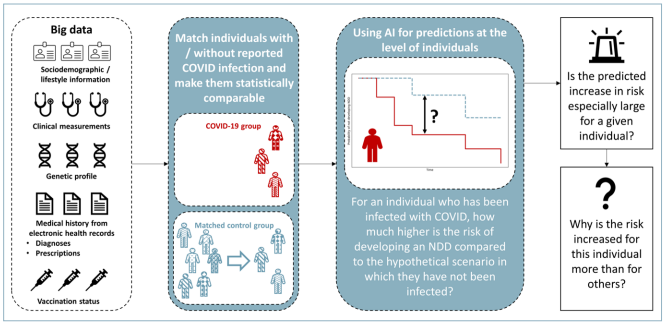

Illustration the approach taken by the Fraunhofer Team led by Prof. Holger Fröhlich:
Big data (including observational data, cohort data, and clinical studies) will be pre-processed and ordered according to exposure to COVID; signals in big data that distinguish the "before COVID" and "after COVID" groups are identified. The research questions and modeling constraints applying to this approach are made clear.

Advanced Data Science and AI for the entire COMMUTE project

WP3 is where all "data-driven" work in COMMUTE is organized. However, WP3 is also the center of gravity when it comes to advanced data analysis for others. All partners in COMMUTE can rely on the combined expertise in WP3 to support them with statistical data analytics and Machine Learning.

Collaboration with other COMMUTE Work Packages

WP3 will closely work with WP2 (the work package where all knowledge is managed and dedicated knowledge graphs are generated that support the functional interpretation of signals generated in WP3) and with WP4, where all molecular and cellular biology experiments and clinical studies are organized.

The expected outcome

WP3 will deliver a set of highly qualified and tested mechanism hypotheses that – if testable in WP4 assays – will be validated in the experimental "wet lab" work package. On the "in silico" side of things, WP3 will develop a set of models predicting individualized risk for COVID-19-induced or COVID-19-enhanced neurodegeneration. A prototypic implementation of a risk app for personalized risk assessment will mark the endpoint of WP3 activities in COMMUTE; we deliberately promise only the prototypic implementation of a risk model ensemble rather than a fully validated health app, which is largely unrealistic.

Contact

Fraunhofer Institute for Algorithms
and Scientific Computing SCAI
Schloss Birlinghoven 1
53757 Sankt Augustin

Prof. Dr. Martin Hofmann-Apitius
Phone +49 2241 14-4103
martin.hofmann-apitius@
scai.fraunhofer.de

www.commute-project.eu