

# Paving the Way towards Personalized Prevention and Care of Severe Norovirus Gastroenteritis (PRESENT)

Sarah Nee, Dominik Wolff, Michael Marschollek

Peter L. Reichertz Institut für Medizinische Informatik der TU Braunschweig und der Medizinischen Hochschule Hannover

## Motivation

Virale Infektionen stellen eine enorme gesundheitliche und ökonomische Belastung dar. Bei vielen dieser Erkrankungen sind die involvierten Pathomechanismen jedoch nicht oder nicht ausreichend bekannt.

Der Großteil der viralen Gastroenteriden wird durch das Norovirus verursacht. Das Virus ist hoch-infektiös und kann bei immungeschwächten Patient\*innen zu schweren Komplikationen führen. Es gibt weder einen Impfstoff gegen das Norovirus, noch eine spezifische antivirale Therapie. Daher ist lediglich eine unterstützende Behandlung der Symptome möglich.

Bekannte Risikofaktoren sind Alter, Geschlecht und Immundefizienz.

Eine erhöhte Resistenz gegenüber dem Norovirus haben Menschen, die eine nicht funktionierende FUT2-Genvariante tragen. Doch auch einige Menschen mit funktionierender Variante scheinen weniger anfällig zu sein. Das Mikrobiom des Darms könnte ebenfalls im Zusammenhang mit der Wirtimmunität stehen. Die komplexen Interaktionen zwischen Virus, Mensch und Mikrobiom erfordern einen holistischen Ansatz.

## Biomarkeridentifikation

Wie bei vielen Erkrankungen spielen beim Verlauf einer Norovirusinfektion verschiedenste Einflussfaktoren, auch Biomarker genannt, eine Rolle.

Um eine personalisierte Behandlung zu ermöglichen müssen sie identifiziert werden. Die dabei zu analysierende Datenmenge ist sehr groß und beschreibt verschiedene Modalitäten. In PRESENT werden klinische Daten zusammen mit biologischen Daten, wie den genetischen Informationen (Genom) und den im Darm vorkommenden Mikroorganismen (Mikrobiom) der Patient\*innen sowie den von ihnen ausgeschiedenen Proteinen (Proteom) berücksichtigt. Mittels der gemeinsamen Analyse dieser Daten können noch nicht bekannte Zusammenhänge aufgedeckt werden.

Projektpartner



Gefördert durch



Hierzu werden Methoden der künstlichen Intelligenz (KI) eingesetzt, welche basierend auf Häufigkeit und Verteilung der Daten automatisiert Gemeinsamkeiten und damit Einflussfaktoren innerhalb einer bestimmten Patientengruppe, beispielsweise Patient\*innen mit einem schweren oder chronischen Gastroenteritisverlauf, identifizieren können. Da seit der COVID-19 Pandemie die Anzahl an Norovirusinfektionen drastisch zurückgegangen ist, werden die Analysen auf in den vergangenen Jahren routinemäßig erhobenen retrospektiven Daten der Medizinischen Hochschule Hannover und anderen Quellen durchgeführt. Dies bringt einige Nachteile mit sich. Einerseits sind die erhobenen Parameter und die Größe der Studienpopulation

nicht wählbar, wie es bei einer vorausschauenden (prospektiven) Studie der Fall wäre. Andererseits unterscheiden sich die für verschiedene Patienten erhobenen Parameter, sodass Lücken im Datensatz vorkommen. Ebenso ist nicht immer eine klare Einteilung der Patient\*innen in Gruppen, beispielsweise basierend auf dem Schweregrad, vorhanden. Um mit diesen Besonderheiten umzugehen, werden spezielle Verfahren der KI eingesetzt. Über ein sogenanntes halb-

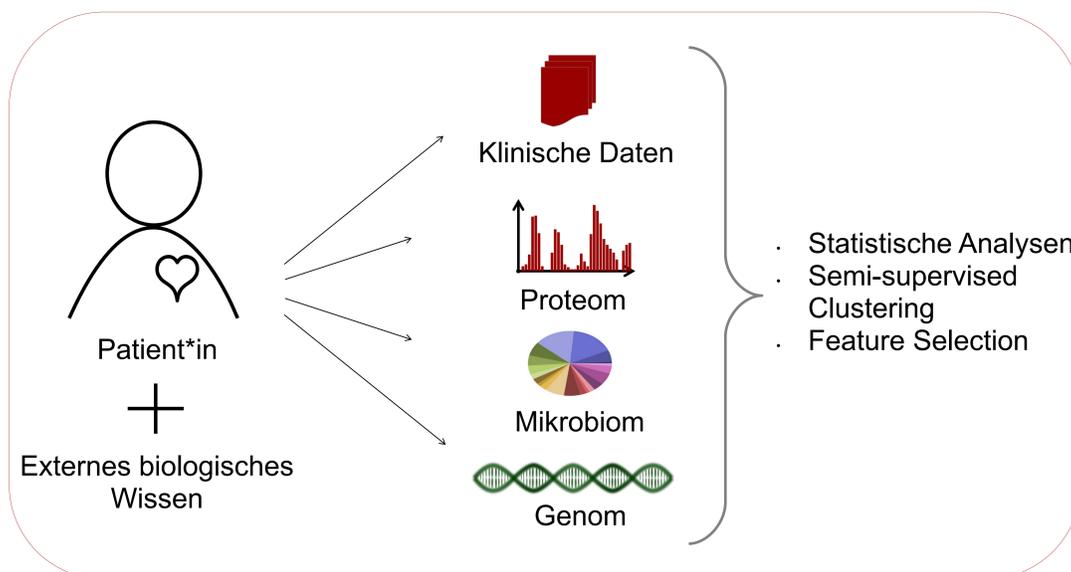


Abbildung 1: Die zu analysierenden Daten mit ihren entsprechenden Methoden.

überwachtes (semi-supervised) Clustering, wird ein maschineller Lernalgorithmus trainiert, der die Daten mit und ohne Gruppeninformationen mithilfe von zusätzlichem (biologischen) Wissen nach ihrer Ähnlichkeit gruppiert.

Sind die Gruppenzugehörigkeiten bekannt, kann ein KI-Modell trainiert werden, welches lernt, zwischen den Gruppen zu unterscheiden. Aus diesem Modell können die Faktoren (Biomarker), die den größten Einfluss auf die Entscheidungsfindung haben abgelesen werden (Feature Selection).

