

Aus der Klinik für Psychiatrie und Psychotherapie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin
Campus Benjamin Franklin

DISSERTATION

**Autismus-Spektrum-Störungen im Jugend- und Erwachsenenalter:
Kann die Diagnostik mit Hilfe von Verfahren des Maschinellen
Lernens weiter optimiert werden?**

**Autism-spectrum-disorders in adolescence and adulthood:
Is it possible to further optimize the diagnostic process using
machine learning?**

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Charlotte Küpper

aus Hannover

Datum der Promotion: 04.06.2021

Inhaltsverzeichnis

1. Abkürzungsverzeichnis.....	3
2. Abstract.....	4
3. Zusammenfassung	6
4. Einführung	8
5. Fragestellung	15
6. Methodik	16
6.1 Studiendesign	16
6.2 Statistische Analysen	20
7. Ergebnisse.....	23
7.1 Ergebnisse in Bezug auf die erste Fragestellung: Gesamtstichprobe	23
7.2 Ergebnisse in Bezug auf die zweite Fragestellung: Alters-Subgruppen	27
8. Diskussion	30
9. Literaturverzeichnis.....	38
10. Eidesstattliche Versicherung	44
11. Anteilserklärung.....	45
12. Auszug aus der Journal Summary List (ISI Web of Knowledge SM).....	47
13. Druckexemplar der ausgewählten Publikation: Küpper, C., Stroth, S., Wolff, N., Hauck, F., Kliewer, N., Schad-Hansjosten, T., Kamp-Becker, I., Poustka, L., Roessner, V., Schultebrucks, K. & Roepke, S. (2020). Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. <i>Scientific Reports</i> 10, 4805.	48
14. Lebenslauf	60
15. Komplette Publikationsliste	63
16. Danksagung.....	64

1. Abkürzungsverzeichnis

ADI-R	Autism Diagnostic Interview Revised (deutsch <i>Diagnostisches Interview für Autismus - Revidiert</i>)
ADOS	Autism Diagnostic Observation Schedule (deutsch <i>Diagnostische Beobachtungsskala für Autistische Störungen</i>)
APA	American Psychological Association
ASD	Autism-Spectrum-Disorders (deutsch <i>Autismus-Spektrum-Störungen</i>)
AQ	Autismus-Spektrums-Quotient
AUC	Area under the Curve (deutsch <i>Fläche unter der Kurve</i>)
BEC	best-estimate clinical diagnosis (deutsch <i>best-mögliche klinische Diagnose</i>)
BMBF	Bundesministerium für Bildung und Forschung
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
ICD-10	International Classification of Diseases and Related Health Problems, 10th Edition
ML	Maschinelles Lernen
ROC	Receiver Operating Characteristic (deutsch <i>Grenzwertoptimierungskurve</i>)
SRS	Social Responsiveness Scale
SVM	Support-Vektor-Maschinen
WHO	Weltgesundheitsorganisation

2. Abstract

Introduction: Autism spectrum disorders (ASD) are serious and lifelong disorders that significantly impair the quality of life of those affected throughout the entire life span. Previous studies show that a diagnosis of high-functioning ASD often goes unrecognized until adolescence and adulthood. However, a correct diagnosis and precise differentiation between co-morbidity and differential diagnosis is of great importance for those affected and for the planning of appropriate psychosocial interventions. Diagnosing ASD in adolescence and adulthood is a complicated and time-consuming process, which requires the use of various standardized diagnostic tools and high clinical expertise. One of the most widely used behavioral diagnostic tools is the Autism Diagnostic Observation Schedule (ADOS). Previous work using machine learning techniques suggested that ASD detection in children can be achieved with substantially fewer items than those comprising the full ADOS (Modules 1-3). Here, we expand on this work with a specific focus on adolescents and adults as assessed with the ADOS Module 4. The present study investigates whether machine learning techniques can be used to identify a reduced subset of psychopathological characteristics from the ADOS Module 4 that can precisely differentiate between adolescents and adults with ASD and adolescents and adults with other clinical disorders.

Methods: A sub-sample of the "ASD-Net" consortium consisting of 673 diagnosed cases from a clinical population was examined. The clinical diagnosis was based on the established international diagnostic "gold standard" of ASD. 57% of all cases received a diagnosis of ASD ("ASD": n=385, age 25.6 years, 74% male, IQ=104.7) and 43% did not receive a diagnosis of ASD but relevant differential diagnoses or no current psychiatric disorders ("non-ASD": n=288, age 26.8 years, 73% male, IQ=104.8). The data was analyzed using Support Vector Machine (SVM) to identify a subset of items from the ADOS module 4 that differentiate between the two classes (ASD vs. non-ASD).

Results: We identified reduced subsets of 5 behavioral features for the whole sample as well as for each age subgroup (adolescents vs. adults) that showed good specificity and sensitivity. Our reduced subsets reached performance comparable to that of the full

ADOS (consisting of 31 items) and depict a substantial reduction in the number of items that have to be coded.

Conclusion: Although all items of the ADOS capture relevant behavioural concepts, the identified behavioural characteristics may include essential constructs that differentiate particularly well between individuals with ASD and individuals with other clinically complex presentations. These results may help to improve the complicated diagnostic process of ASD by encouraging future efforts to develop novel diagnostic instruments for ASD detection based on the identified constructs as well as aiding clinicians in the difficult question of differential diagnosis.

3. Zusammenfassung

Einleitung: Autismus-Spektrum-Störungen (*autism spectrum disorders*, ASD) sind schwerwiegende und lebenslange Erkrankungen, die die Lebensqualität der Betroffenen in allen Lebensbereichen beeinträchtigen. Studien zeigen, dass die Diagnose einer hochfunktionalen ASD oftmals bis ins Jugend- und Erwachsenenalter unerkannt bleibt. Im fortgeschrittenen Alter ist eine Diagnosestellung häufig durch das Fehlen verlässlicher Aussagen über die frühe Entwicklung sowie durch Symptomüberlappungen verschiedener Störungsbilder mit ASD erschwert. Eine korrekte Diagnosestellung ist jedoch für die Betroffenen und für die Planung geeigneter psychosozialer Interventionen von hoher Bedeutung. Die Diagnosestellung einer ASD ist ein komplizierter und zeitaufwendiger Prozess, der den Einsatz verschiedener Diagnoseinstrumente erfordert. Eines der am häufigsten verwendeten Diagnoseinstrumente ist die Diagnostische Beobachtungsskala für Autistische Störungen (ADOS). Frühere Studien konnten mithilfe von Verfahren des maschinellen Lernens zeigen, dass die ASD-Klassifizierung bei Kindern (ADOS Module 1-3) mit wesentlich weniger Items erreicht werden kann als mit der gesamten ADOS. In der vorliegenden Studie wird diese Fragestellung auf das Jugend- und Erwachsenenalter (ADOS Modul 4) ausgeweitet. Es wird untersucht, ob mithilfe von Verfahren des maschinellen Lernens Merkmale aus der ADOS identifiziert werden können, die gut zwischen Jugendlichen und Erwachsenen mit ASD und mit anderen klinischen Störungsbildern differenzieren.

Methodik: Untersucht wurden 673 diagnostizierte Fälle einer psychiatrischen Inanspruchnahme-Population. Die klinische Diagnosestellung erfolgte nach etabliertem Goldstandard. Bei 57% der Fälle lag eine ASD Diagnose vor („ASD“: n=385, Alter 25,6 Jahre, 74% männlich, IQ=104,7), bei 43% lagen relevante Differentialdiagnosen oder keine psychischen Störungen vor („non-ASD“: n=288, Alter 26,8 Jahre, 73% männlich, IQ=104,8). Die Daten wurden mittels Support Vector Machine (SVM) analysiert.

Ergebnisse: Es ließen sich reduzierte Subsets bestehend aus 5 Verhaltensmerkmalen für die Gesamtstichprobe sowie für die Subgruppen (Jugendliche vs. Erwachsene) identifizieren, die eine gute Spezifität und Sensitivität aufwiesen. Die identifizierten Subsets stellen eine bedeutsame Reduzierung der zu kodierenden Items dar und

wiesen dabei vergleichbar hohe Klassifizierungsleistungen auf wie die gesamte ADOS (bestehend aus 31 Items).

Schlussfolgerung: Obwohl alle Items der ADOS relevante Verhaltenskonzepte umfassen, erfassen die identifizierten Items möglicherweise wesentliche Merkmale, die besonders gut zwischen Personen mit ASD und Personen mit anderen klinischen Störungsbildern differenzieren. Die Ergebnisse können einen Beitrag zur Verbesserung des ASD-Diagnoseprozesses bei Jugendlichen und Erwachsenen leisten, indem die identifizierten Items die Grundlage für die Entwicklung neuer Screening-Instrumente liefern und Kliniker*innen bei der schwierigen diagnostischen Entscheidung unterstützen.

4. Einführung

Autismus-Spektrum-Störungen (engl. *autism-spectrum-disorders*, kurz ASD¹) umfassen eine Gruppe von tiefgreifenden Entwicklungsstörungen mit einer Prävalenz von etwa 1% in der Gesamtbevölkerung (Lai et al., 2014). ASD sind gekennzeichnet durch früh einsetzende, anhaltende und qualitative Beeinträchtigungen und Auffälligkeiten in zwei Kernbereichen: erstens im Bereich der wechselseitigen sozialen Kommunikation und Interaktion, sowie zweitens im Bereich eingeschränkter, sich wiederholender, stereotyper Verhaltensweisen oder Interessen (APA, 2013). Diese qualitativen Beeinträchtigungen liegen von Geburt an vor und stellen funktions- und situationsübergreifende Merkmale dar (*tiefgreifende* Entwicklungsstörungen). Der Begriff „Spektrum“ bezieht sich auf den großen Umfang an Symptomen und Fähigkeiten sowie auf das weitreichende Ausmaß an Beeinträchtigungen, die bei Personen mit ASD vorhanden sein können. In der in Deutschland gegenwärtig noch geltenden 10. Ausgabe der Internationalen Klassifikation psychischer Störungen der Weltgesundheitsorganisation (WHO; *International Classification of Diseases and Related Health Problems* (ICD-10), 2016) wird zwischen drei autistischen Störungsbildern unterschieden: dem Frühkindlichen Autismus (ICD-10: F84.0), dem Atypischen Autismus (ICD-10: F84.1) und dem Asperger-Syndrom (ICD-10: F84.5). Da sich jedoch eine qualitative Abgrenzung zwischen diesen unterschiedlichen Varianten sowohl theoretisch als auch praktisch als schwierig und wenig trennscharf erwiesen hat, und sich lediglich graduelle Unterschiede zwischen den verschiedenen Störungsbildern aufweisen lassen, wurden in der Neuauflage des amerikanischen Klassifikationssystems der American Psychological Association (APA; *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), 2013) sämtliche autistische Störungen unter dem Sammel-Begriff „Autismus-Spektrum-Störung“ subsummiert (Lai et al., 2014). Auch die Neuauflage des ICD, das ICD-11, welche im Januar 2022 in Kraft treten soll, hat diese Entwicklung mit aufgenommen.

In den letzten Jahrzehnten ist das Bewusstsein für ASD in der Öffentlichkeit als auch in der Gesundheitsversorgung deutlich gestiegen. Dies zeigt sich unter anderem in einem bemerkenswerten Anstieg der Prävalenzzahlen über die letzten Jahrzehnte (ca. 0,04%

¹ Im Folgenden wird die geläufige englische Abkürzung ASD verwendet.

in den 1970er Jahren im Vergleich zu ca. 1% aktuell; für einen Überblick siehe Lai et al., 2014), der hauptsächlich auf die Erweiterung der diagnostischen Kriterien sowie eine bessere gesellschaftliche Aufklärung zurückgeführt wird. Hierdurch ist jedoch auch die Inanspruchnahme von und der Bedarf an diagnostischen Untersuchungen und Behandlungen von ASD durch qualifizierte Kliniker*innen deutlich angestiegen.

Die Diagnosestellung einer ASD ist ein komplizierter, langwieriger und kostenaufwendiger Prozess, der einen hohen Grad an klinischer Expertise erfordert (Kamp-Becker et al., 2018). Obwohl die Forschung ständig Fortschritte im Bereich der zugrunde liegenden genetischen und neurobiologischen Faktoren macht, die mit ASD assoziiert sind, und es starke Evidenz dafür gibt, dass ASD genetisch determinierte Störungsbilder sind, gibt es derzeit keine zuverlässigen biologischen Marker für ASD (Lai et al., 2014). Die Diagnosestellung einer ASD basiert nach wie vor überwiegend auf der klinischen Beobachtung und (fremd-)anamnestischen Beschreibung von Verhaltenssymptomen (Hayes et al., 2018). Aufgrund der beträchtlichen Heterogenität in der Ausprägung und Schwere der Kern- und Begleitsymptomatik ist der Einsatz verschiedener psychodiagnostischer Instrumente für ein klinisches Urteil unerlässlich. Der derzeitige sogenannte „internationale Goldstandard“ der ASD-Diagnostik umfasst den Einsatz verschiedener standardisierter Selbst- und Fremdbeurteilungsfragebögen sowie klinischer Diagnoseinstrumente, die den Kliniker*innen helfen sollen, eine bestmögliche klinische Diagnose zu stellen (die sogenannte *best-estimate clinical diagnosis*, BEC; Wigham et al., 2018). Zwei der am weitesten verbreiteten klinischen Diagnoseinstrumente sind die Diagnostische Beobachtungsskala für Autistische Störungen (ADOS bzw. ADOS-2 für die revidierte zweite Auflage; Lord et al., 2000; Lord et al., 2012) und das diagnostische Interview für Autismus (ADI-R; Rutter et al., 2003).

Das ADI-R ist ein halb-strukturiertes Interview, das fremdanamnestisch mit Eltern oder anderen engen Bezugspersonen durchgeführt wird und sich auf die Entwicklungsgeschichte und das Vorliegen von autistischen Symptomen in der Kindheit konzentriert. Die Befragung erfasst Symptome aus den Bereichen soziale Interaktion, Kommunikation und repetitive Verhaltensweisen und bezieht dabei den Entwicklungsverlauf der zu diagnostizierenden Person mit ein. Da für eine ASD Diagnose entscheidende Symptome bereits in früher Kindheit vorgelegen haben müssen, ist eine Erfassung von Symptomen aus diesem Zeitraum unerlässlich. Die Auswertung erfolgt anhand eines empirisch generierten diagnostischen Algorithmus.

Die ADOS ist eine standardisierte, halbstrukturierte diagnostische Beobachtungsskala, die dazu dient, wichtige sozial-kommunikative Verhaltensweisen sowie stereotype und sich wiederholende Verhaltensmerkmale gezielt hervorzurufen und zu beurteilen. Hierzu erschafft der/die Diagnostiker*in durch sein/ihr eigenes Verhalten gezielt soziale Situationen, die einen eindeutigen Aufforderungscharakter gegenüber den Proband*innen aufweisen und kodiert anschließend das Auftreten oder Ausbleiben spezifischer Verhaltensweisen auf verschiedenen Items. Diese Items lassen sich verschiedenen Domänen zuordnen: „Kommunikation“, „Soziale Interaktion“, „Vorstellungsvermögen“, „Stereotype Verhaltensweisen und Eingeschränkte Interessen“ und „Andere Auffällige Verhaltensweisen“. Die ADOS umfasst verschiedene Untersuchungs-Module (ADOS Modul 1-4), die je nach Alter und expressivem Sprachniveau der Proband*innen zum Einsatz kommen. Für jedes Modul gibt es einen diagnostischen Algorithmus (der sogenannte „ADOS-Algorithmus“), der aus ausgewählten Items besteht, welche den DSM-IV und ICD-10 ASD-Kriterien entsprechen und sich als besonders reliabel und valide erwiesen haben (für eine detaillierte Beschreibung der Entwicklung der ADOS-Algorithmen siehe Lord et al., 2000 und Hus & Lord, 2014). Die Kodierungen dieser ausgewählten Items werden summiert und mit Cut-Off-Werten verglichen, um eine Instrument-Klassifizierung von „ASD“ oder „kein ASD“ und somit Rückschlüsse auf die zugrundeliegende Diagnose zu ermöglichen. Das Modul 4 ist für verbal fließend sprechende Jugendliche und Erwachsene konzipiert und umfasst den Einsatz von 15 Aufgaben und 31 zu kodierenden Items. Die durchschnittliche Untersuchungsdauer liegt bei 40-60 Minuten und einer anschließenden Kodierungsdauer von bis zu 30 Minuten. Bei Modul 4 fließt eine Auswahl von 11 Items aus den Domänen „Soziale Interaktion“ und „Kommunikation“ in den ADOS-Algorithmus ein, der die Instrument-Klassifizierung von „ASD“ oder „kein ASD“ ermöglicht.

Die ADOS ist eines der psychometrisch am besten evaluierten diagnostischen Instrumente bei ASD, insbesondere bei Kindern (Module 1 und 2; Le Couteur et al., 2008; Kamp-Becker et al., 2018), mit insgesamt weniger Studien im Bereich des Jugend- und Erwachsenenalters (Module 3 und 4; Langmann et al., 2017). Obwohl allgemein eine gute psychometrische Qualität der ADOS in Forschungskontexten nachgewiesen wurde, deutet eine Vielzahl von Studien auf einen geringeren diagnostischen Nutzen in der klinischen Praxis hin (Wigham et al., 2018; Zander et al.,

2016). So konnten Lord und Kolleg*innen (2000) in ihrer ursprünglichen Publikation zur ADOS beispielsweise zeigen, dass das Modul 4 gut zwischen Personen mit ASD und gesunden Kontroll-Personen trennt, mit hohen Sensitivitäts- und Spezifitätswerten von >90%. Sensitivität (die richtig-positiv-Rate eines Instruments) und Spezifität (die richtig-negativ-Rate eines Instruments) sind eng miteinander verbunden und beide gleichermaßen wichtig im diagnostischen Prozess, wobei das Ziel der Untersuchung bestimmt, welches Maß noch wichtiger ist: Eine hohe Sensitivität ist unerlässlich, wenn das Ziel der Untersuchung ist auf ASD zu screenen, um keinen potenziellen Fall zu verpassen. Eine hohe Spezifität hingegen ist entscheidend, wenn es darum geht, sicher zu sein, dass bei einer Person tatsächlich eine ASD vorliegt. In Bezug auf naturalistische klinische Settings (d.h. den Vergleich von Personen mit ASD und Personen mit anderen klinischen Störungsbildern) zeigte die ADOS Modul 4 insgesamt deutlich geringere Sensitivitäts- als auch Spezifitätswerte, mit starken Schwankungen in Abhängigkeit der untersuchten klinischen Vergleichsstichprobe (Werte zwischen 50% und 90%; für einen aktuellen Überblick siehe Wigham et al., 2018). Beeinträchtigungen in der sozialen Interaktion und Kommunikation können bei vielen verschiedenen psychischen Störungen auftreten, so dass das Risiko einer falsch-positiven Klassifizierung mit der ADOS erhöht ist (Maddox et al., 2017). Zudem weisen Erwachsene mit einer hochfunktionalen ASD oftmals effektive Kompensationsstrategien auf, welche Defizite im Bereich Sozialer Interaktion und Kommunikation überdecken können (das sogenannte „*Camouflaging*“; Lai et al., 2017). Dies kann in der Folge dazu führen, dass die zugrundeliegenden Defizite im Rahmen der ADOS Untersuchung nicht entdeckt werden, wodurch das Risiko einer falsch-negativen Klassifizierung erhöht wird. Eine korrekte Abgrenzung von ASD zu anderen Störungsbildern ist jedoch insbesondere im klinischen Alltag und für die Planung geeigneter Behandlungen und Unterstützungsmaßnahmen von hoher Relevanz.

Obwohl die meisten ASD-Fälle in der Kindheit diagnostiziert werden, bleibt die Diagnose dennoch in einigen Fällen bis zum Jugend- und Erwachsenenalter unbemerkt, insbesondere bei Personen mit durchschnittlichen bis überdurchschnittlichen kognitiven und sprachlichen Fähigkeiten, ausgeprägten adaptiven Funktionen und subtileren Symptomdarstellungen („hochfunktionale ASD“; Lai & Baron-Cohen, 2015; Happé et al., 2016). Die Diagnosestellung einer ASD bei hochfunktionalen Jugendlichen und Erwachsenen ist im Vergleich zum Kindesalter

zusätzlich aufgrund verschiedener Faktoren erschwert: oftmals sind Personen, die eine verlässliche Aussage über die Entwicklungsgeschichte und frühe Kindheit tätigen können (wie mit dem ADI-R Interview erhoben) nicht verfügbar (Lai & Baron-Cohen, 2015), und die Selbstbeschreibung des Individuums in Bezug auf Symptome kann aufgrund einer Beeinträchtigung der selbst-referenziellen Kognition vermindert sein (Lombardo et al., 2007). Darüber hinaus können erlernte Kompensationsstrategien Beeinträchtigungen verdecken, wodurch die diagnostische Genauigkeit von Beobachtungsinstrumenten wie der ADOS vermindert wird (Lai & Baron-Cohen, 2015). Weiterhin weisen die meisten Erwachsenen mit ASD mindestens eine komorbide psychiatrische Störung auf (Joshi et al., 2013; Lai et al., 2014) und die Symptome von ASD überschneiden sich häufig mit denen anderer psychiatrischer Erkrankungen (Maddox et al., 2017), was die Differentialdiagnostik weiter erschwert. Zudem kommt es insbesondere bei hochfunktionalen Personen, bei denen der Verdacht einer ASD erstmals im fortgeschrittenen Alter geäußert wird, nicht selten vor, dass die Ergebnisse der eingesetzten Diagnose-Instrumente (z.B. ADOS, ADI-R) nicht übereinstimmen. Darüber hinaus existieren Deutschlandweit nur wenige Diagnostikzentren, die auf die ASD-Diagnostik im Jugend- und Erwachsenenalter spezialisiert sind, so dass die Nachfrage bei weitem den Bedarf übersteigt und es für Betroffene zu langen Wartezeiten von einigen Monaten bis Jahren kommen kann (Kamp-Becker et al., 2017). Eine korrekte und zeitnahe Diagnosestellung und eine präzise Unterscheidung zwischen Komorbidität und Differentialdiagnose ist jedoch für die Betroffenen sowie für die Einleitung effizienter Behandlungen und die Planung geeigneter psychosozialer Interventionen von großer Bedeutung (Tromans et al., 2018).

Um den Diagnoseprozess zu optimieren, haben Wissenschaftler*innen in den letzten Jahren begonnen, Verfahren des maschinellen Lernens (engl. *machine learning*, kurz ML) im Bereich der ASD-Diagnostik anzuwenden. Das maschinelle Lernen ist ein multidisziplinäres Forschungsgebiet, das Informatik, Mathematik, künstliche Intelligenz, Suchmethoden und andere Wissenschaften integriert, um zugrunde liegende Muster und genaue Vorhersagemodelle aus bestehenden Datensätzen abzuleiten. ML-Methoden (wie z.B. neuronale Netze, Support-Vektor-Maschinen (SVM), Entscheidungsbäume oder regelbasierte Klassifikatoren) zeichnen sich dadurch aus, dass sie automatisierte Ansätze sind, die wenig formale Annahmen treffen, nur minimale menschliche Beteiligung während der Datenverarbeitung erfordern und zudem

die Fähigkeit aufweisen, Strukturen und Muster aus umfangreichen Datensätzen zu extrahieren (Bzdok & Meyer-Lindenberg, 2018). In der Medizin werden ML-Methoden in erster Linie dazu eingesetzt, um automatisierte Ansätze sowohl für die Diagnose als auch für die Behandlung spezifischer Störungen zu generieren. Ziel der ML-Verfahren ist es, auf der Grundlage bestehender Datenbanken, die vielfältige Merkmale (engl. *features*) enthalten, krankheitsspezifische Muster bzw. Merkmalssubsets zu identifizieren und daraus klinische Vorhersagen zu entwickeln und Entscheidungen zu treffen. Die aus den Daten gewonnenen Erkenntnisse können dann verallgemeinert und für neue Problemlösungen verwendet werden.

Im Bereich der ASD-Diagnostik können ML-Verfahren beispielsweise dafür genutzt werden, um die aussagekräftigsten Merkmale eines bestehenden Diagnose-Instruments zu identifizieren und die Anzahl benötigter Items für eine Klassifizierung (ASD vs. kein ASD) zu reduzieren. Bisherige Studien im Bereich der ASD-Diagnostik haben in Hinblick auf diese Fragestellungen eine Vielzahl von existierenden diagnostischen Instrumenten mittels ML-Methoden untersucht, über Selbst- und Fremdbeurteilungsfragebögen bis zu standardisierten klinischen Diagnoseinstrumenten wie ADOS und ADI-R (für einen aktuellen Überblick siehe Thabtah (2018)).

In Bezug auf existierende Screening-Fragebögen wie den Autismus-Spektrums-Quotienten (AQ; Thabtah & Peebles, 2019) oder die Social Responsiveness Scale (SRS; Bone et al., 2016) konnte beispielsweise durch ML-Verfahren eine reduzierte Anzahl zentraler Items identifiziert werden, welche eine vergleichbar hohe Genauigkeit wie das eigentliche Instrument aufwiesen. Hierdurch ließ sich die Effizienz und Zugänglichkeit dieser Fragebögen deutlich erhöhen.

Ähnliche Experimente wurden durchgeführt, um eine reduzierte Anzahl von Verhaltensmerkmalen aus klinischen Diagnoseinstrumenten wie ADOS Modul 1 bis 3 (Bone et al., 2015; Kosmicki et al., 2015; Levy et al., 2017; Wall et al., 2012) und ADI-R (Bone et al., 2015; Bone et al., 2016; Wall et al., 2012) zu identifizieren. Kosmicki und Kolleg*innen (2015) konnten beispielsweise mittels Support-Vektor-Maschine (SVM) ein reduziertes Subset bestehend aus 12 Items aus ADOS Modul 3 (welches gesamt 28 Items enthält) identifizieren. Dieses 12-Item Subset wies eine hohe Prädiktionsleistung auf (Sensitivitäts- und Spezifitätswerte von 98% bzw. 97%) und stellte dabei eine Halbierung der zu kodierenden Items dar (57%ige Reduzierung). Vergleichbare Subsets

bestehend aus minimal 5 bis maximal 12 Items, die ähnlich hohe Vorhersagewerte aufwiesen (Sensitivität und Spezifität >90%), ließen sich für das ADOS Modul 1 (Bone et al., 2015; Wall et al., 2012), Modul 2 (Kosmicki et al., 2015; Levy et al., 2017) und Modul 3 (Levy et al., 2017) identifizieren. Die Ergebnisse dieser Studien deuten darauf hin, dass die Erkennung von ASD bei Kindern mit wesentlich weniger Items im Vergleich zu den ursprünglichen ADOS- und ADI-R-Algorithmen erreicht werden kann, wobei eine hohe diagnostische Genauigkeit, Sensitivität und Spezifität erhalten bleibt. Die Autor*innen schlussfolgern, dass sich durch reduzierte und verbesserte Algorithmen die zeit- und kostenaufwendige ASD-Diagnostik optimieren ließe. Einschränkend ist jedoch zu sagen, dass die Aussagekraft und Generalisierbarkeit dieser Studien durch verschiedene Faktoren vermindert ist. So sind die meisten der verwendeten Datensätze unausgewogen und bestehen aus sehr vielen ASD und wenigen non-ASD Fällen (z.B. 90% ASD vs. 10% non-ASD bei Kosmicki et al., 2015; vgl. auch Bone et al., 2015), oder fokussieren sich auch hier nur auf die Unterscheidung von ASD-Fällen und gesunden Kontroll-Personen (vgl. Wall et al., 2012). Zudem stammen die meisten der Datensätze aus Archivbeständen (vgl. Bone et al., 2015; Kosmicki et al., 2015; Levy et al., 2017; Wall et al., 2012) und die ASD-Diagnose basiert teilweise nur auf der ADOS-Klassifikation und entspricht somit nicht einer *best-estimate* klinischen Diagnose (vgl. Kosmicki et al., 2015; Levy et al., 2017). Nichtsdestotrotz sind diese Ergebnisse vielversprechend und unterstützen die Hypothese, dass es möglich ist, den komplizierten und zeitaufwendigen ASD-Diagnose-Prozess mittels ML-Verfahren weiter zu optimieren. Dieser Ansatz soll auch in der vorliegenden Arbeit verfolgt werden. Die Identifizierung einer reduzierten Anzahl wesentlicher Verhaltensmerkmale, die präzise zwischen ASD- und non-ASD-Fällen unterscheidet, könnte auf vielfältige Weise zu einer Verbesserung des komplexen diagnostischen Prozesses beitragen: erstens zur Optimierung bestehender diagnostischer Instrumente (d.h. zur Überarbeitung existierender Klassifikationsalgorithmen), zweitens zur Verkürzung der Durchführungszeit bestehender diagnostischer Instrumente und/oder drittens zur Entwicklung neuer diagnostischer Instrumente und Methoden für das Erst-Screening auf der Grundlage dieser wesentlichen Merkmale.

Die bisherigen Untersuchungen in diesem Bereich haben sich ausschließlich auf Kinder und jüngere Jugendliche fokussiert (erfasst mittels ADOS Modul 1 bis 3). Unseres Wissens nach existiert bisher keine Studie, die diese Fragestellung in Bezug auf ältere

Jugendliche und Erwachsene (erfasst mittels ADOS Modul 4) untersucht hat. Ziel der vorliegenden Arbeit ist es daher, die vorhandene Literatur mit einem spezifischen Fokus auf hochfunktionale Jugendliche und Erwachsene zu erweitern. Obwohl sich alle Items des ADOS Modul 4 auf relevante Verhaltenskonzepte fokussieren, ist es denkbar, dass einige Items eine höhere Diskriminations- und Klassifizierungsfähigkeit aufweisen, insbesondere in einer komplexen klinischen Stichprobe von Jugendlichen und Erwachsenen. Die in der vorliegenden Arbeit untersuchte große (N=673) und ausgewogene (57% ASD vs. 43% non-ASD) klinische Inanspruchnahmepopulation von Jugendlichen und Erwachsenen mit *best-estimate* klinischen Diagnosen, welche alle mit dem initialen Verdacht auf das Vorliegen einer ASD vorstellig wurden, stellt zudem einen Vorzug im Vergleich zu früheren Studien in diesem Bereich dar.

5. Fragestellung

In der vorliegenden Arbeit werden die folgenden zwei Fragestellungen untersucht:

1. Kann mithilfe von Verfahren des maschinellen Lernens eine reduzierte Anzahl psychopathologischer Merkmale aus ADOS Modul 4 identifiziert werden, die präzise zwischen Jugendlichen und Erwachsenen mit ASD und Jugendlichen und Erwachsenen mit anderen klinischen Störungsbildern differenziert?
2. Finden sich Unterschiede in den identifizierten Merkmalen zwischen Jugendlichen (≤ 21 Jahre) und Erwachsenen (> 21 Jahre)?

6. Methodik

6.1 Studiendesign

Die vorliegende Studie wurde als Multi-Center Studie im Rahmen des „ASD-Net“-Konsortiums (Kamp-Becker et al., 2017) realisiert und vom Bundesministerium für Bildung und Forschung (BMBF) gefördert. Teilnehmende Studienzentren waren vier spezialisierte ASD-Ambulanzen in Deutschland (Charité Universitätsmedizin Berlin, Philipps-Universität Marburg, Technische Universität Dresden und Zentralinstitut für Seelische Gesundheit Mannheim), in denen der diagnostische „Goldstandard“ der ASD-Diagnostik angewendet wird, um die Diagnose einer ASD zu bestätigen oder auszuschließen.

Im Rahmen des ASD-Net Projekts wurden klinische Routine-Diagnostik-Daten aus diesen vier Studienzentren in eine gemeinsame Datenbank eingegeben und retrospektiv anonymisiert ausgewertet (retrospektive Akten-Analyse, engl. *retrospective chart review*). In die Datenbank eingeschlossen wurden die Daten aller Kleinkinder, Kinder, Jugendlichen und Erwachsenen, die zwischen 2001 und 2015 mit dem Verdacht auf das Vorliegen einer ASD als Teil der gängigen klinischen Praxis standardisiert untersucht wurden (klinische Inanspruchnahmepopulation bestehend aus N=2453 Personen, Altersspanne: 1-72 Jahre). Alle Patient*innen wurden von Fachärzt*innen oder Psychotherapeut*innen an die Spezialambulanzen überwiesen oder wurden selbständig vorstellig. In der vorliegenden Studie wurde eine Teilstichprobe dieser „ASD-Net“ Gesamtstichprobe bestehend aus n=673 diagnostizierten Jugendlichen und Erwachsenen untersucht.

Dieses Vorgehen wurde von der Ethik-Kommission der Charité – Universitätsmedizin Berlin mit der Registrierungsnummer (EA4/129/19) genehmigt.

Diagnostische Verfahren

In allen vier Studienzentren wurde der sogenannte „internationale Goldstandard“ der ASD-Diagnostik angewandt, um die Diagnose einer ASD gesichert zu vergeben oder auszuschließen. Das diagnostische Vorgehen umfasste eine standardisierte Verhaltensbeobachtung (ADOS Modul 4; Lord et al., 2000), ein standardisiertes Eltern-Interview sofern durchführbar (ADI-R; Rutter et al., 2003) sowie eine umfassende differentialdiagnostische Untersuchung mittels in Deutschland etablierten Strukturierten

Klinischen Interviews und Fragebögen. Die Erhebung der standardisierten Verhaltensbeobachtung mittels ADOS Modul 4 und die differentialdiagnostische Untersuchung war in allen 673 Fällen der untersuchten Teil-Stichprobe möglich, das standardisierte Eltern-Interview konnte nur in 62% aller Fälle erhoben werden (ASD: 71%, non-ASD: 50%).

ADOS

Die ADOS ist eine standardisierte, diagnostische Beobachtungsskala, mit der wichtige sozial-kommunikative Verhaltensweisen sowie stereotype und repetitive Verhaltensmerkmale erfasst werden können (Lord et al., 2000). In Abhängigkeit von Alter und expressivem Sprachniveau kann eins von mehreren Module ausgewählt werden, um anhand von gezielt inszenierten Aufgaben, Aktivitäten und Gesprächen relevante autistische Symptome zu prüfen. Das Modul 4 ist für fließend sprechende ältere Jugendliche und Erwachsene konzipiert. Die Durchführung der Verhaltensbeobachtung dauert im Schnitt 40 bis 60 Minuten. Im Anschluss folgt eine Auswertung anhand validierter Kodierrichtlinien auf 31 verschiedenen Items, welche sich den Domänen „Kommunikation“, „Soziale Interaktion“, „Vorstellungsvermögen“, „Stereotype Verhaltensweisen und Eingeschränkte Interessen“ und „Andere Auffällige Verhaltensweisen“ zuordnen lassen. Die Bewertung dieser Items erfolgt anhand bestimmter Codes, die pro Item vergeben werden können. Die Codes liegen auf einer Ordinal-Skala, die von 0 („Kein Anhaltspunkt für eine Auffälligkeit im Sinne einer autistischen Störung“) über 1 („Das Verhalten ist etwas auffällig bzw. merkwürdig, aber nicht unbedingt eindeutig abnorm“), 2 („Das Verhalten ist eindeutig abnorm, für eine autistische Störung typisch“) bis hin zu 3 („Das Verhalten ist schwer abnorm“) reicht. Zudem ist es möglich zusätzliche Codes von 7 („Das Verhalten ist abnorm, aber nicht im Sinne des Autismus“) und 8 („Logisch nicht beurteilbar, da das Verhalten bspw. während der Beobachtung nicht gezeigt wurde“) zu vergeben. Fehlende Werte (d.h. ausgelassene Aufgaben oder leer gelassene Itemkodierungen) werden mit einer 9 kodiert.

Eine Auswahl der kodierten Items wird in einem diagnostischen Algorithmus verrechnet (der sogenannte „ADOS-Algorithmus“). Dieser Bewertungsalgorithmus besteht aus 11 Items (siehe Tabelle 3) aus den Domänen „Soziale Interaktion“ und „Kommunikation“ und ermöglicht die Berechnung eines Vergleichswertes, anhand dessen eine ADOS-Klassifizierung von „Autismus“ oder „Nicht-Autismus“ erfolgen kann.

Stichprobe

Unsere Stichprobe umfasste Daten von insgesamt 673 Fällen. 57% aller Fälle erhielten nach Durchführung der ausführlichen diagnostischen Untersuchung eine ASD Diagnose („ASD“, n=385) und 43% erhielten keine Diagnose aus dem Autismus-Spektrum („non-ASD“, n=288) aber relevante Differentialdiagnosen wie bspw. affektive Störungen, Angststörungen, ADHS und/oder Persönlichkeitsstörungen oder keine aktuelle psychiatrische Diagnose. Eine detaillierte Beschreibung der psychopathologischen Auffälligkeiten unserer Stichprobe ist in Tabelle 1 zu finden.

Tabelle 1. Psychopathologische Beschreibung der Stichprobe ^a

Psychopathologische Beschreibung	ASD (n=385)	non-ASD (n=288)
Andere psychiatrische Störungen außer ASD	n=160 (42%)	n=189 (66%)
ICD-10: F10-F19	n=8 (2%)	n=6 (2%)
<i>Psychische und Verhaltensstörungen durch psychotrope Substanzen</i>		
ICD-10: F20-F29	n=2 (0.6%)	n=3 (1%)
<i>Schizophrenie, schizotype und wahnhaftige Störungen</i>		
ICD-10: F30-F39	n=88 (23%)	n=53 (18%)
<i>Affektive Störungen</i>		
ICD-10: F40-F48	n=56 (15%)	n=61 (21%)
<i>Neurotische, Belastungs- und somatoforme Störungen</i>		
ICD-10: F50-F59	n=2 (0.6%)	n=0 (0%)
<i>Verhaltensauffälligkeiten mit körperlichen Störungen und Faktoren</i>		
ICD-10: F60-F69	n=15 (4%)	n=45 (16%)
<i>Persönlichkeits- und Verhaltensstörungen</i>		
ICD-10: F80-F89 <u>ausgenommen</u> F84	n=15 (4%)	n=17 (6%)
<i>Entwicklungsstörungen</i>		
ICD-10: F90-F98	n=44 (11%)	n=66 (23%)
<i>Verhaltens- und emotionale Störungen mit Beginn in der Kindheit und Jugend</i>		
Keine andere aktuelle psychiatrische Störung	n=225 (58%)	n=99 (34%)

^a Tabelle modifiziert aus: Küpper et al. (2020).

Die ASD-Subtypen gemäß ICD-10 (Autismus (F84.0), Atypischer Autismus (F84.1), Asperger-Syndrom (F84.5)) wurden für die Analysen zusammengruppiert, so dass ein binäres Maß für unsere ML-Analysen erstellt wurde („ASD“ und „non-ASD“). Zwischen den beiden Gruppen fanden sich keine signifikanten Unterschiede hinsichtlich Alter, Geschlecht und IQ (Tabelle 2).

Tabelle 2. Stichprobenbeschreibung ^a

Merkmal	ASD (n=385)	non-ASD (n=288)	Statistischer Test
Durchschnittsalter (SD)	25.63 Jahre (11.27)	26.81 Jahre (12.45)	n.s. (t(582.92)=1.27, p=0.21)
% Alter ≥18 Jahre [n]	67% [n=258]	68% [n=196]	n.s. ($\chi^2(1)=0.082$, p= 0.78)
% Alter >21 Jahre [n]	52.7% [n=203]	51.7% [n=149]	n.s. ($\chi^2(1)=0.065$, p= 0.80)
Geschlecht: % männlich [n]	74.3% männlich [n=286]	72.9% männlich [n=210]	n.s. ($\chi^2(1)=0.16$, p= 0.69)
Mittlerer IQ (SD)*	104.68 (16.00) (basierend auf n=343)	104.84 (15.49) (basierend auf n=245)	n.s. (t(586)=0.12, p=0.90)

^a Tabelle modifiziert aus: Küpper et al. (2020).

Abkürzung: n.s., nicht signifikant; SD, Standardabweichung (englisch *standard deviation*).

*Für 87% der Stichprobe waren vollständige IQ-Daten verfügbar.

Insgesamt zeigte sich in unserer Stichprobe eine große Altersspanne, welche von 10 bis 72 Jahren reichte, mit einem Median von 22 Jahren. Das Grenzalter der Kinder- und Jugendpsychiatrie im Vergleich zur Erwachsenenpsychiatrie liegt in Deutschland bei 21 Jahren, so dass wir unsere Gesamtstichprobe für die Untersuchung der zweiten Fragestellung in die folgenden zwei Untergruppen aufgespaltet haben:

„Jugendliche“ im Alter von 10 bis 21 Jahren (n=321, 56.7% ASD (n=182), Durchschnittsalter=16.42 Jahren (SD=2.14), 81.3% männlich) und „Erwachsene“ im Alter von 22 bis 72 Jahren (n=352, 57.7% ASD (n=203), Durchschnittsalter=34.97 Jahren (SD=9.90), 66.8% männlich).

6.2 Statistische Analysen

In früheren ML-Klassifikationsexperimenten im Bereich der ASD-Diagnostik wurden verschiedene Techniken des maschinellen Lernens eingesetzt, darunter Support-Vektor-Maschinen (SVM), Entscheidungsbäume und allgemeine lineare Modelle (für einen Überblick siehe Thabtah, 2018). In diesen früheren Studien erzielten SVM die besten Ergebnisse (Bone et al., 2016; Kosmicki et al., 2015; Levy et al., 2017). Zudem gehören SVM aufgrund ihrer hohen Vorhersagekraft zu den am häufigsten verwendeten Algorithmen für die ASD-Klassifikation (Thabtah, 2018). Aus diesen Gründen haben wir uns entschieden, SVM als unseren ML-Klassifikator zu verwenden (unter Verwendung von *svmRadial* des *Caret R*-Pakets; Kuhn, 2008). Alle 31 ADOS-Items wurden als Merkmale (engl. *features*) verwendet und die klinische Diagnose (*best-estimate clinical diagnosis*) der Patient*innen wurde als unsere Zielvariable (ASD vs. non-ASD) verwendet. Alle Schritte der Dateninspektion, Datenvorverarbeitung und Datenanalyse wurden mit dem Programm *R* (Version 3.5.1) in *Rstudio* 1.1.456 durchgeführt.

Vorverarbeitung der Daten

Für die Datenvorverarbeitung für unsere ML-Verfahren wurden die ADOS-Codes analog den Kodierrichtlinien des ADOS-Manuals wie folgt umkodiert: Werte von 3 wurden zu 2 und Werte von 7 und 8 wurden zu 0 umkodiert. Fehlende Werte (d.h. ADOS-Codes von 9) wurden mithilfe des *k-nearest-neighbor* Imputationsverfahrens unter Verwendung der *preprocess()*-Funktion aus dem *Caret R*-Paket mit *k=5 nearest neighbors (knnImpute;* Kuhn, 2008) ersetzt. In diesem Verfahren wird ein fehlender Wert durch den Mittelwert aus den *k* Werten nächstgelegener Spalten ersetzt. In unserem Datensatz fehlten sechs Items (Items A6, B4, C1, E1, E2, E3) in bis zu 10% aller Fälle. Alle anderen Items enthielten in weniger als 2.5% der Fälle fehlende Werte. Darüber hinaus wurden alle numerischen Variablen auf den Bereich [0;1] normalisiert.

Um die Wahrscheinlichkeit unvoreingenommener Ergebnisse zu maximieren, wurden strenge Vorkehrungen getroffen, um eine sogenannte Überanpassung unserer Modelle an die Daten (engl. *overfitting*) zu vermeiden. Hierzu wurde der Gesamtdatensatz zunächst zufällig in zwei Partitionen aufgeteilt (75% und 25%). Die zufällige Aufteilung in diese beiden Partitionen wurde für die Zielvariable geschichtet (stratifizierte Zufallsstichprobe). Die 75%ige Partition diente als Trainingsset für die Hyperparameteroptimierung, die Modellbildung und das Modelltraining. Hyperparameter

stellen Parameter dar, die vor Beginn des Modellbildungs- und Modelltrainingsprozesses festgelegt werden müssen und so abgestimmt werden, dass das Modell die Klassifikationsaufgabe optimal lösen kann. Für den Modellbildungsprozess (Merkmalsauswahl) und das Modelltraining wurde in beiden Fällen eine 5-fach wiederholte 10-fache Kreuzvalidierung (engl. *5-times repeated 10-fold cross validation*) angewandt, d.h. das Trainingsset wurde zufällig in $k=10$ gleich große Teilmengen aufgeteilt. Neun ($k-1$) dieser Teilmengen wurden zum Erstellen und Trainieren der Modelle verwendet und eine Teilmenge wurde zum Validieren verwendet. Dieser Kreuzvalidierungsprozess wurde 10-mal wiederholt (die sogenannten „Faltungen“, engl. *folds*), wobei jede der Teilmengen genau einmal als Validierungsmenge verwendet wurde. Die Ergebnisse aus den 10 Faltungen wurden anschließend gemittelt, um eine Schätzung zu erhalten. Dieses Kreuzvalidierungsverfahren wurde insgesamt 5-mal wiederholt und die fünf Ergebnisse wurden erneut gemittelt, um eine einzige Schätzung zu erhalten. Bei jedem Kreuzvalidierungsschritt wurde die Vorverarbeitung (Wertebereichs-Normalisierung und Imputation fehlender Werte) erneut für jede Faltung durchgeführt.

Die 25%ige Partition diente als Testset für die Evaluation der Vorhersagekraft der endgültigen Modelle in völlig „ungesehenen“ Fällen, d.h. Fälle, die nicht für die Erstellung der Modelle verwendet wurden.

Zur Evaluation der Modelleleistungen wurden *Receiver Operating Characteristic* (ROC)-Kurven und die Fläche unter der ROC-Kurve (engl. *Area under the ROC-Curve*, AUC) verwendet. Eine ROC-Kurve ist eine grafische Abbildung der Vorhersageleistung der Klassifizierungsmodelle und bildet zwei Parameter ab: auf der Y-Achse wird die richtig-positiv-Rate (d.h. Sensitivität) und auf der X-Achse die falsch-positiv-Rate (d.h. 1-Spezifität) abgetragen. Die Fläche unter der ROC-Kurve (AUC) misst die gesamte zweidimensionale Fläche unter der Kurve und sagt aus, inwieweit das Modell in der Lage ist, korrekt zwischen den Klassen (ASD vs. non-ASD) zu unterscheiden. Die Werte der AUC reichen von 0 bis 1; je höher der AUC-Wert, desto besser ist das Modell in der Lage ASD-Fälle (richtig-positiv-Rate) und non-ASD-Fälle (richtig-negativ-Rate) jeweils als solche vorherzusagen. Ein Wert von 0,5 würde eine zufällige Klassifikation darstellen und ein Wert $<0,5$ würde darauf hinweisen, dass das Modell die Klassen falsch klassifiziert.

Datenanalyse mittels Machine Learning

Unser ML-Ansatz bestand aus mehreren Schritten: Zunächst wurde eine Merkmalsauswahl (engl. *feature selection*) anhand des Trainingssets durchgeführt, um eine reduzierte Anzahl von Merkmalen mit hohen prädiktiven Leistungsprofilen zu identifizieren. Hierfür wurde eine rekursive Merkmalsauswahl (engl. *recursive feature selection, rFuncs*) unter Anwendung von *Random Forests* (Breiman, 2001) über das *Caret R*-Paket (Kuhn, 2008) durchgeführt. Als Ergebnismaß für die Hyperparameteroptimierung für unsere Merkmalsauswahl wurde der Kappa-Koeffizient von Cohen verwendet, wobei Sensitivität und Spezifität gleichermaßen gewichtet wurden, und es wurde eine 5-fach wiederholte 10-fache Kreuzvalidierung angewandt.

Der zweite Schritt bestand darin, unser reduziertes Modell auf dem Trainingsset zu trainieren. Auch während des Modelltrainings wurde eine 5-fach wiederholte 10-fache Kreuzvalidierung angewandt, um stabile Vorhersageergebnisse zu erzielen. Als Ergebnismaß für die Hyperparameteroptimierung wurde die Fläche unter der ROC-Kurve (AUC) verwendet. Die abgestimmten Hyperparameter waren *Sigma* und *Cost*, welche mittels Ansatz der Zufallsuche (engl. *random search*) aus 500 verschiedenen Kombinationen bestimmt wurden. Alle anderen abstimmbaren Parameter wurden auf Standardwerte gesetzt (Karatzoglou et al., 2004).

Der dritte Schritt umfasste das Testen unseres reduzierten Modells auf dem dezidierten Testset, um zu sehen, wie gut das reduzierte Modell ASD Fälle von non-ASD Fällen in völlig „ungesehenen“ Daten vorhersagen kann. Als Ergebnismaß wurde auch hier die Fläche unter der ROC-Kurve (AUC) verwendet.

In einem letzten Schritt wurden die Modelleleistungen miteinander verglichen. Hierzu verglichen wir den AUC-Wert unseres neu-identifizierten reduzierten Modells mit dem AUC-Wert des bestehenden 11-Item ADOS-Algorithmus und dem AUC-Wert aller 31 ADOS-Items. Zudem verglichen wir die Vorhersageleistung unseres reduzierten Modells mit dem AUC-Wert des 12-Item-Modells, das von Kosmicki und Kolleg*innen (2015) in einem früheren Experiment anhand von ADOS Modul 3 Daten (Kinder und Jugendliche) identifiziert wurde (eine Liste dieser 12 Items findet sich in Tabelle 3). Die Unterschiede wurden mittels DeLong-Test (DeLong et al., 1988) sowie Bootstrap-Resampling-Signifikanztest (Carpenter & Bithell, 2000; Pepe et al., 2009) für zwei korrelierte ROC-Kurven quantifiziert (wobei jedes Mal die Überlappung der Konfidenzintervalle mit 10000 Bootstrap-Iterationen verglichen wurde).

Alle der oben genannten Schritte wurden zunächst in der gesamten Stichprobe („alle Altersgruppen“, N=673) und anschließend in den Alters-Subgruppen („Jugendliche“, n=321 und „Erwachsene“, n=352) durchgeführt.

7. Ergebnisse

7.1 Ergebnisse in Bezug auf die erste Fragestellung: Gesamtstichprobe

Kann mithilfe von Verfahren des maschinellen Lernens eine reduzierte Anzahl psychopathologischer Merkmale aus ADOS Modul 4 identifiziert werden, die präzise zwischen Jugendlichen und Erwachsenen mit ASD und Jugendlichen und Erwachsenen mit anderen klinischen Störungsbildern differenziert?

Durch den rekursiven Algorithmus zur Merkmalsauswahl wurden fünf Merkmale als die wichtigsten ausgewählt, d.h. diejenigen Merkmale, die im Durchschnitt die höchste Fähigkeit zur Vorhersage von Personen mit ASD im Vergleich zu Personen mit anderen klinischen Störungsbildern während der Kreuzvalidierung aufwiesen:

- A9 (Deskriptive, konventionelle, instrumentelle oder informative Gesten),
- B1 (Ungewöhnlicher Blickkontakt),
- B2 (An andere gerichtete Gesichtsausdrücke),
- B10 (Qualität der sozialen Reaktion) und
- B11 (Umfang der wechselseitigen sozialen Kommunikation).

Alle fünf ausgewählten Merkmale stammen aus den Domänen „Soziale Interaktion“ und „Kommunikation“ der ADOS und sind sowohl im ursprünglichen ADOS-Algorithmus (bestehend aus 11 Merkmalen) als auch in dem von Kosmicki und Kolleg*innen (2015) vorgeschlagenen 12-Item-Modell enthalten (siehe Tabelle 3). Die Items A9, B1 und B2 stellen Auffälligkeiten in der reziproken nonverbalen Kommunikation und Interaktion der Proband*innen dar, die während der ADOS-Untersuchung beobachtet werden. Die Items B10 und B11 stellen qualitative und zusammenfassende Items dar, bei denen die Kliniker*innen Auffälligkeiten im gesamten sozialen Verhalten der Proband*innen während der ADOS-Untersuchung bewerten.

Tabelle 3. Die 11 Merkmale aus dem ADOS-Modul-4-Algorithmus und die 5 Merkmale, die durch den Merkmalsauswahlprozess für die gesamte Stichprobe identifiziert wurden (fett gedruckt)^a

Code	Merkmalsbeschreibung	ADOS Domäne
A4*	Stereotypisierter / Idiosynkratischer Gebrauch von Wörtern oder Phrasen	Kommunikation
A8*	Konversation	Kommunikation
A9*	Deskriptive, konventionelle, instrumentelle oder informative Gesten	Kommunikation
A10	Emphatische oder Emotionale Gesten	Kommunikation
B1*	Ungewöhnlicher Blickkontakt	Reziproke Soziale Interaktion
B2*	An andere gerichtete Gesichtsausdrücke	Reziproke Soziale Interaktion
B6	Empathie	Reziproke Soziale Interaktion
B8	Eigenverantwortlichkeit	Reziproke Soziale Interaktion
B9*	Qualität der sozialen Annäherungsversuche	Reziproke Soziale Interaktion
B10*	Qualität der sozialen Reaktion	Reziproke Soziale Interaktion
B11*	Umfang der wechselseitigen sozialen Kommunikation	Reziproke Soziale Interaktion

^a Tabelle modifiziert aus: Küpper et al. (2020)

* Items, die auch in dem von Kosmicki und Kolleg*innen (2015) identifizierten 12-Item-Modell enthalten sind. Weitere Items, die von Kosmicki et al. identifiziert wurden und die nicht im ADOS-Algorithmus enthalten sind, sind: A7 (Bericht über Ereignisse), D1 (ungewöhnliches sensorisches Interesse an Spielmaterial/Person), D2 (Hand-, Finger- und andere komplexe Manierismen) und D4 (übermäßiges Interesse an ungewöhnlichen oder hochgradig spezifischen Themen oder Objekten).

Unser reduziertes Modell bestehend aus 5 Merkmalen erreichte einen AUC-Wert von 0.87 im Trainingsset, was vergleichbar ist mit der Leistung des 11-Item-Modells des ADOS-Algorithmus, des 31-Item-Modells unter Verwendung aller ADOS-Items, und des 12-Item-Modells von Kosmicki und Kolleg*innen (2015) (alle AUC-Werte von 0.87; siehe Tabelle 4).

Zur unabhängigen Validierung wurde die Leistung der Modelle auf dem Testset berechnet. Bei der unabhängigen Vorhersage erreichten alle Modelle etwas niedrigere AUC-Werte (siehe Tabelle 4): Unser reduziertes Modell mit 5 Merkmalen erreichte einen AUC-Wert von 0.82 (Sensitivität = 0.71, Spezifität = 0.83) im Vergleich zu AUC-Werten von 0.84 des 11-Item-Modells (Sensitivität = 0.85, Spezifität = 0.76), des von Kosmicki et al. (2015) vorgeschlagenen 12-Item-Modells (Sensitivität = 0.77, Spezifität = 0.82) und des 31-Item-Modells (Sensitivität = 0.79, Spezifität = 0.81; siehe Tabelle 4). Für eine grafische Darstellung der Ergebnisse in Form von ROC-Kurven siehe Abb. 1.

Tabelle 4. Vorhersageleistung der Modelle für das Trainings- und Testset in Bezug auf die gesamte Stichprobe („Alle Altersgruppen“) ^a

		SVM Modelle			
		5-Items*	11-Items (ADOS Algorithmus)	31-Items (Alle ADOS Items)	12-Items (Kosmicki et al., 2015)
Alle Altersgruppen (n=673)					
Trainingsset					
AUC (Sensitivität, Spezifität)	0.87 (0.72, 0.87)	0.87 (0.75, 0.88)	0.87 (0.73, 0.88)	0.87 (0.73, 0.85)	
Testset					
AUC (Sensitivität, Spezifität)	0.82 (0.71, 0.83)	0.84 (0.85, 0.76)	0.84 (0.79, 0.81)	0.84 (0.77, 0.82)	
*Items A9, B1, B2, B10, B11					

^a Tabelle modifiziert aus: Küpper et al. (2020)

In einem letzten Schritt wurden die Leistungen der Modelle miteinander verglichen. Es fanden sich keine signifikanten Unterschiede zwischen den AUC-Werten unseres reduzierten 5-Item-Modells und dem 11-Item-Modell des ADOS-Algorithmus (DeLong's Test: $Z=-1,63$, $p=0,10$; Bootstrapping²: $D=-1,61$, $p=0,11$), dem 12-Item-Modell von Kosmicki et al. (2015) (DeLong's Test: $Z=-1,27$, $p=0,20$; Bootstrapping²: $D=-1,26$, $p=0,21$) und dem 31-Item-Modell (DeLong's Test: $Z=-1,29$, $p=0,20$; Bootstrapping²: $D=-1,26$, $p=0,21$).

² Anzahl der neu gesampelten Bootstrap-Proben jeweils = 10000

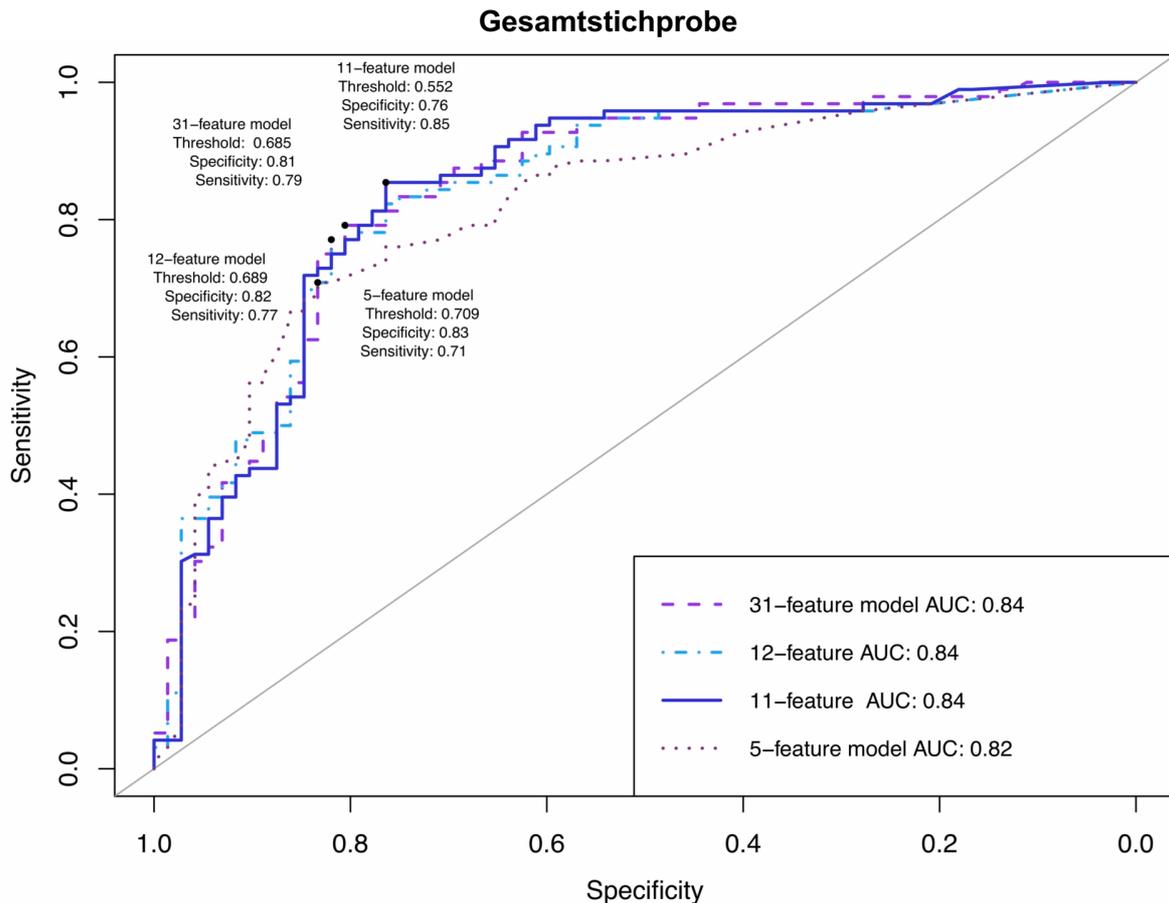


Abbildung 1. ROC-Kurven der vier Modelle in Bezug auf das Testset für die gesamte Stichprobe („alle Altersgruppen“). Die Y-Achse stellt die richtig-positiv-Rate (Sensitivität) und die X-Achse die falsch-positiv-Rate dar (1-Spezifität; zur verbesserten Interpretierbarkeit wurde der Wert in der Grafik invertiert). Die diagonale Linie stellt die Leistung des Zufallsprozesses dar (engl. *chance line*). Der optimale ROC-Schwellenwert ist mit der höchsten Summe von Sensitivität und Spezifität gekennzeichnet (schwarze Punkte; Robin et al., 2011).^a
^a Abbildung entnommen aus: Küpper et al., 2020.

Zusammengefasst weisen die Ergebnisse zur ersten Fragestellung darauf hin, dass sich mittels SVM ein reduziertes Subset bestehend aus 5 Merkmalen aus ADOS Modul 4 identifizieren ließ, das vergleichbar gut zwischen Jugendlichen und Erwachsenen mit ASD und Jugendlichen und Erwachsenen mit anderen klinischen Störungsbildern differenziert wie der bestehende ADOS-Algorithmus und die gesamte ADOS bestehend aus 31 Items. Im Vergleich zur gesamten ADOS stellt dieses reduzierte Subset jedoch eine deutliche Reduzierung der zu kodierenden Items dar (Reduzierung um 84%).

7.2 Ergebnisse in Bezug auf die zweite Fragestellung: Alters-Subgruppen

Finden sich Unterschiede in den identifizierten Merkmalen zwischen Jugendlichen (≤ 21 Jahre) und Erwachsenen (> 21 Jahre)?

Im Vergleich zu unserem reduzierten 5-Item-Modell der gesamten Stichprobe („alle Altersgruppen“: Items A9, B1, B2, B10, B11) wählte unser rekursiver Merkmalsauswahl-Algorithmus für die spezifischen Alters-Subgruppen leicht unterschiedliche Merkmale als die wichtigsten aus: Items A9, B1, B2, B3, B9 für „Jugendliche“ und Items A9, B2, B3, B9, B10 für „Erwachsene“. Ähnlich wie bei der Gesamtstichprobe stammen jedoch alle ausgewählten Merkmale aus den ADOS-Domänen „Soziale Interaktion“ und „Kommunikation“. Zwei Items (B3: Sprachproduktion und verbundene nonverbale Kommunikation und B9: Qualität der sozialen Annäherungen) wurden in beiden Alters-Subgruppen, aber nicht in der Gesamtstichprobe ausgewählt. Item B9 stellt ein qualitatives und zusammenfassendes Item dar, das die Gesamtqualität der Versuche des Individuums bewertet, soziale Interaktionen mit dem/der Diagnostiker*in zu initiieren. Item B3 enthält Informationen darüber, wie die Vokalisierungen des/der Proband*in mit den nonverbalen Verhaltensweisen (wie Blickkontakt, Gestik und Mimik) koordiniert sind.

Durch die Aufspaltung der Stichprobe in Alters-Subgruppen konnten wir die Genauigkeit unserer Modelle im Vergleich zur Gesamtstichprobe weiter erhöhen.

In der Subgruppe der Erwachsenen erreichte das identifizierte 5-Item-Modell einen AUC-Wert von 0.84 im Testset, und war damit numerisch marginal geringer als die AUC-Werte der 11-Item- und 31-Item-Modelle (beide AUCs = 0.87) und der AUC-Wert des von Kosmicki et al. (2015) vorgeschlagenen 12-Item-Modells (AUC = 0.85; siehe Tabelle 5). In der Subgruppe der Jugendlichen erreichte das identifizierte 5-Item-Modell im Testset einen AUC-Wert von 0.90 und lag damit numerisch leicht über den Werten des 11-Item-Modells (AUC = 0.88), des 31-Item-Modells (AUC = 0.87) und des 12-Item-Modells (AUC = 0.84; siehe Tabelle 5).

Vergleichbar mit der Gesamtstichprobe gab es auch hier keine signifikanten Unterschiede zwischen den AUC-Werten der 5-Item-Modelle und dem vom ADOS-Algorithmus vorgeschlagenen 11-Item-Modell („Jugendliche“: DeLong’s test: $Z=0.85$,

$p=.40$; bootstrapping: $D=0.85$, $p=.39$; „Erwachsene“: DeLong’s test: $Z=-1.60$, $p=.11$; bootstrapping: $D=-1.54$, $p=.12$), dem 31-Item-Modell („Jugendliche“: DeLong’s test: $Z=0.94$, $p=.35$; bootstrapping: $D=0.93$, $p=.35$; „Erwachsene“: DeLong’s test: $Z=-1.76$, $p=.08$; bootstrapping: $D=-1.71$, $p=.09$) und dem von Kosmicki et al. (2015) vorgeschlagenen 12-Item-Modell („Jugendliche“: DeLong’s test: $Z=1.49$, $p=.14$; bootstrapping: $D=1.48$, $p=.14$; „Erwachsene“: DeLong’s test: $Z=-0.30$, $p=.76$; bootstrapping: $D=-0.30$, $p=.76$) in den Testsets der jeweiligen Subgruppen.

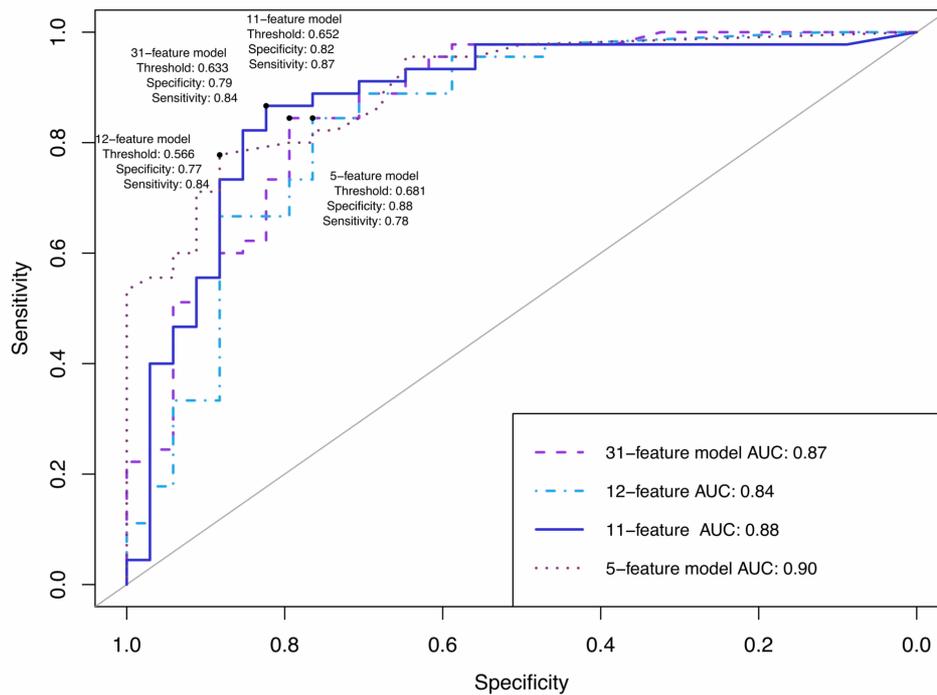
Eine grafische Darstellung der ROC-Kurven für die jeweiligen Alters-Subgruppen findet sich in Abbildung 2.

Tabelle 5. Vorhersageleistung der Modelle für das Trainings- und Testset in Bezug auf die Alters-Subgruppen („Jugendliche“ und „Erwachsene“) ^a

	SVM Modelle			
	5-Items*	11-Items (ADOS Algorithmus)	31-Items (alle ADOS Items)	12-Items (Kosmicki et al., 2015)
Jugendliche ≤ 21 Jahre (n=321)				
Trainingsset				
AUC (Sensitivität, Spezifität)	0.83 (0.67, 0.85)	0.85 (0.58, 0.92)	0.84 (0.66, 0.85)	0.85 (0.70, 0.86)
Testset				
AUC (Sensitivität, Spezifität)	0.90 (0.78, 0.88)	0.88 (0.87, 0.82)	0.87 (0.84, 0.79)	0.84 (0.84, 0.77)
*Items: A9, B1, B2, B3, B9				
Erwachsene > 21 Jahre (n=352)				
Trainingsset				
AUC (Sensitivität, Spezifität)	0.87 (0.69, 0.88)	0.88 (0.71, 0.89)	0.86 (0.65, 0.89)	0.86 (0.62, 0.92)
Testset				
AUC (Sensitivität, Spezifität)	0.84 (0.90, 0.76)	0.87 (0.92, 0.84)	0.87 (0.90, 0.84)	0.85 (0.90, 0.78)
*Items: A9, B2, B3, B9, B10				

^a Tabelle modifiziert aus: Küpper et al. (2020)

Jugendliche



Erwachsene

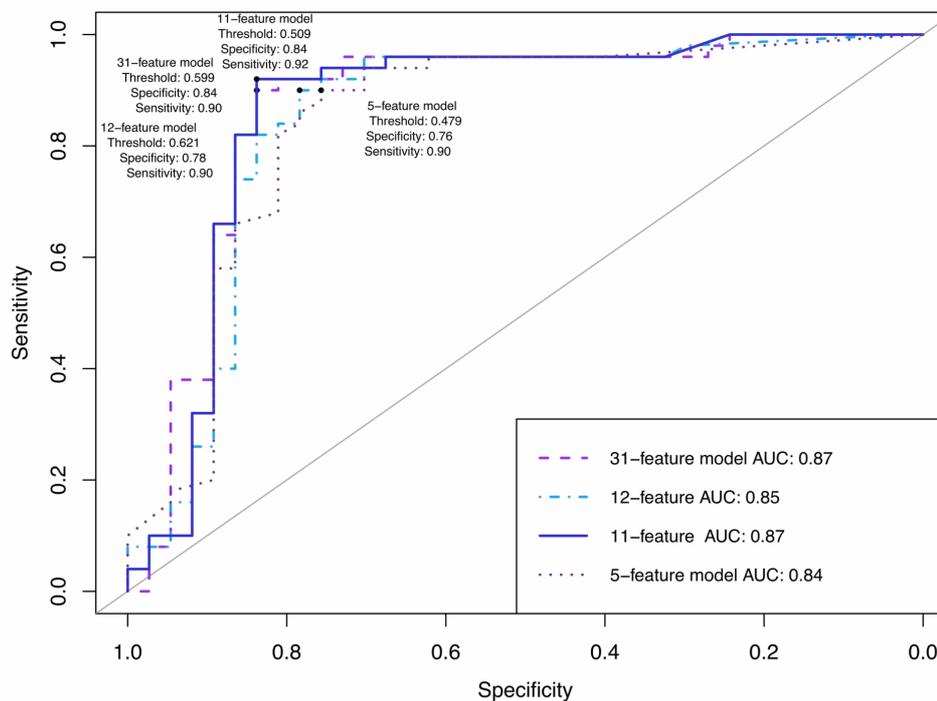


Abbildung 2. ROC-Kurven der vier Modelle in Bezug auf das Testset für die Alters-Subgruppen („Jugendliche“ und „Erwachsene“). Die Y-Achse stellt die richtig-positiv-Rate (Sensitivität) und die X-Achse die falsch-positiv-Rate dar (1-Spezifität; zur verbesserten Interpretierbarkeit wurde der Wert in der Grafik invertiert). Die diagonale Linie stellt die Leistung des Zufallsprozesses dar (engl. *chance line*). Der optimale ROC-Schwellenwert ist mit der höchsten Summe von Sensitivität und Spezifität gekennzeichnet (schwarze Punkte; Robin et al., 2011).^a

^a Abbildungen entnommen aus: Küpper et al., 2020.

Zusammengefasst weisen die Ergebnisse zur zweiten Fragestellung darauf hin, dass es leichte Unterschiede in den identifizierten Merkmalen zwischen Erwachsenen (>21 Jahre) und Jugendlichen (≤ 21 Jahre) gab. Obwohl alle ausgewählten Items den beiden Domänen „Soziale Interaktion“ und „Kommunikation“ entsprangen, zeigten sich leichte Unterschiede in der Zusammensetzung der ausgewählten Merkmale in den jeweiligen Subgruppen. Durch eine Aufspaltung in Alters-Subgruppen ließ sich zudem die Vorhersagegenauigkeit der reduzierten Modelle im Vergleich zur Gesamtstichprobe noch weiter erhöhen.

8. Diskussion

In der vorliegenden Arbeit wurde erstmals die Fragestellung untersucht, ob sich mit Hilfe von Methoden des maschinellen Lernens eine reduzierte Anzahl wesentlicher Verhaltensmerkmale aus ADOS Modul 4 identifizieren lassen, die präzise zwischen Jugendlichen und Erwachsenen mit ASD und Jugendlichen und Erwachsenen mit anderen klinischen Störungsbildern differenziert. Zudem wurde untersucht, ob sich Unterschiede in den identifizierten Merkmalen zwischen Jugendlichen (≤ 21 Jahre) und Erwachsenen (>21 Jahre) finden.

Unter Anwendung eines SVM-basierten ML-Ansatzes konnten wir eine reduzierte Anzahl von fünf Verhaltensmerkmalen aus ADOS Modul 4 identifizieren, die eine gute Spezifität (83%) und Sensitivität (71%) sowie eine insgesamt hohe Klassifizierungsleistung (AUC von 82%) aufwies. Die Vorhersageleistung dieses reduzierten Algorithmus war zudem vergleichbar mit der Leistung des ADOS-Algorithmus (bestehend aus 11 Items), der gesamten ADOS (bestehend aus 31 Items) und einem zuvor von Kosmicki und Kolleg*innen (2015) identifizierten 12-Item-Algorithmus (alle drei AUCs von 84%) und es ließen sich keine signifikanten Unterschiede in der Vorhersageleistung finden.

Durch eine Aufspaltung unserer Gesamtstichprobe in Alters-Subgruppen von Jugendlichen (≤ 21 Jahre) und Erwachsenen (>21 Jahre) ließ sich die Vorhersagegenauigkeit der reduzierten 5-Item-Algorithmen noch weiter erhöhen („Jugendliche“: AUC von 90%; „Erwachsene“: AUC von 84%). Ähnlich wie in der gesamten Stichprobe („alle Altersgruppen“) zeigten sich auch hier keine signifikanten

Unterschiede beim Vergleich der Vorhersageleistungen der verschiedenen Modelle in den jeweiligen Subgruppen („Jugendliche“: AUCs von 90% (5-Items) vs. 88% (11-Items) vs. 87% (31-Items) vs. 84% (12-Items, vorgeschlagen von Kosmicki et al. (2015)); „Erwachsene“: AUCs von 84% (5-Items) vs. 87% (11-Items) vs. 87% (31-Items) vs. 85% (12-Items, vorgeschlagen von Kosmicki et al. (2015))).

Die von uns berichteten Sensitivitäts- und Spezifitätswerte waren insgesamt geringer als die Werte, die in früheren Arbeiten zu Modul 1 bis 3 berichtet wurden (Sensitivitäts- und Spezifitätswerte von >90%; Bone et al., 2015; Kosmicki et al., 2015; Levy et al., 2017; Wall et al., 2012). Diese Differenz zeigte sich sowohl in Bezug auf die von uns identifizierten reduzierten 5-Item-Subsets als auch die gesamte ADOS und den 11-Item ADOS Algorithmus und könnte verschiedene Ursachen haben: zum einen fokussiert sich die Mehrzahl dieser früheren Arbeiten auf die Unterscheidung von ASD-Fällen vs. gesunden Kontroll-Personen, wodurch sich die hohen Sensitivitäts- und Spezifitätswerte erklären ließen. Unsere Stichprobe hingegen bestand aus einer komplexen klinischen Inanspruchnahmepopulation, in der alle Proband*innen initial den Verdacht auf das Vorliegen einer ASD hatten und vielfältige weitere Auffälligkeiten der sozialen Interaktion aufwiesen. Zum anderen wurden in der vorliegenden Arbeit erstmalig ältere Jugendliche und Erwachsene untersucht, wohingegen sich frühere Arbeiten ausschließlich auf das Kindes- und frühe Jugendalter fokussiert haben. Die Unterscheidung zwischen ASD-Fällen vs. Personen mit anderen psychischen Auffälligkeiten kann insbesondere im Erwachsenenalter durch das bewusste Überdecken (erhöhtes Risiko einer falsch-negativen Klassifizierung aufgrund von „*Camouflaging*“) oder die Überlappung von Symptomen (erhöhtes Risiko einer falsch-positiven Klassifizierung) erschwert sein. Obwohl die von uns berichteten Werte insgesamt geringer sind, so fallen sie dennoch mit Sensitivitäts- und Spezifitätswerten zwischen 70% und 90% in einen guten diagnostischen Bereich. Zudem sind unsere Ergebnisse im Einklang mit den Befunden aus früheren Arbeiten, die vergleichbare Vorhersageleistungen der reduzierten Subsets und des ADOS-Algorithmus bzw. der gesamten ADOS berichten.

Bei genauerer Betrachtung der ausgewählten Items zeigte sich, dass alle Items unserer reduzierten Algorithmen (für die Gesamtstichprobe als auch die Alters-Subgruppen) aus den beiden ADOS Domänen „Soziale Interaktion“ und „Kommunikation“ stammen. Alle

fünf Items, die für die gesamte Stichprobe identifiziert wurden („alle Altersgruppen“: A9, B1, B2, B10, B11), sind auch im ursprünglichen 11-Item ADOS-Algorithmus enthalten. Dieses Ergebnis bestätigt einerseits die prädiktive Güte des ADOS-Algorithmus, könnte jedoch auch darauf hinweisen, dass im bestehenden 11-Item-Algorithmus redundante diagnostische Informationen enthalten sind. Für die Alters-Subgruppen wurden verglichen mit der Gesamtstichprobe leicht unterschiedliche Items ausgewählt. Interessanterweise zeigte sich eine starke Überlappung der ausgewählten Items über beide Subgruppen hinweg („Jugendliche“: A9, B1, B2, B3, B9; „Erwachsene“: A9, B2, B3, B9, B10). Dieses Ergebnis könnte auf die hohe Bedeutung dieser Items (A9, B2, B3, B9) unabhängig vom Entwicklungsstand der Betroffenen hinweisen. Den einzigen Unterschied zwischen den beiden Alters-Subgruppen stellte die Identifikation des „basaleren“ sozial-kommunikativen Items B1 (Ungewöhnlicher Blickkontakt) in der Gruppe der Jugendlichen und stattdessen die Identifikation des „komplexeren“ qualitativen und zusammenfassenden Items B10 (Qualität der sozialen Reaktion) in der Gruppe der Erwachsenen dar. Dieses Ergebnis legt nahe, dass Jugendliche im Vergleich zu Erwachsenen mit ASD einen auffälligeren Blickkontakt zeigen, während Erwachsene eher „komplexere“ Beeinträchtigungen der sozialen Interaktion aufweisen. Dieser Befund ist im Einklang mit früheren Studien, die darauf hindeuten, dass sich die Darstellung von ASD-Symptomen im Laufe des Lebens verändert aufgrund erlernter kompensatorischer Verhaltensweisen und genereller Entwicklungsfortschritte (Howlin et al., 2013; LeCouteur et al., 2008).

Vergleicht man unsere identifizierten Items mit denen früherer ML-Experimente (Kosmicki et al., 2015; Levy et al., 2017; Wall et al., 2012) sowie noch unveröffentlichter Ergebnisse unserer Arbeitsgruppe zum ADOS Modul 1 bis 3, so finden sich auch hier deutliche Parallelen in den ausgewählten Merkmalen über die verschiedenen Studien und Altersgruppen hinweg. Diese Parallelen zeigen sich insbesondere bei den Merkmalen B1, B2, B9, B10 und B11, die alle aus der ADOS Domäne „Soziale Interaktion“ stammen. Obwohl alle Items der ADOS relevante Verhaltenskonzepte in Bezug auf die soziale Interaktion erfassen, sind Beeinträchtigungen in der sozialen Interaktion nicht ASD-spezifisch, sondern treten bei vielen verschiedenen psychiatrischen Störungen auf. Die von uns identifizierten Items könnten jedoch relevante „Kern“-Merkmale der sozialen Interaktion darstellen, die spezifischer für Jugendliche und Erwachsene mit ASD sind und eine gute Diskriminationsfähigkeit zu

anderen klinischen Störungsbildern aufweisen. Eine Fokussierung auf solche Items, die besonders spezifisch für ASD sind und weniger ausgeprägt bei Personen mit anderen klinischen Störungsbildern auftreten, könnte Kliniker*innen in dem schwierigen Prozess der differentialdiagnostischen Entscheidungsfindung helfen.

Unsere Ergebnisse deuten darauf hin, dass eine reduzierte Anzahl von nur fünf Items eine ebenso präzise Vorhersage von ASD ermöglicht wie der vom ADOS-Manual vorgeschlagene 11-Item-Algorithmus oder sogar die gesamte ADOS bestehend aus 31 Items. Im Vergleich zur gesamten ADOS stellen unsere identifizierten 5-Item-Algorithmen jedoch eine erhebliche Reduzierung in der Anzahl der Verhaltensmerkmale auf, die für die Beurteilung kodiert werden müssen (eine Reduzierung von insgesamt 84% in Bezug auf die gesamte ADOS und 54% in Bezug auf den ADOS-Algorithmus). Bone und Kolleg*innen (2015) vertreten jedoch die Position, dass die Durchführungszeit der ADOS nicht einfach durch eine Reduzierung der zu kodierenden Items minimiert werden kann, da die einzelnen ADOS-Items nicht direkt an eine der 15 Aufgaben der ADOS gebunden sind. Um die Items zu kodieren, sei es immer notwendig, die ADOS-Untersuchung in ihrer ganzen Länge (d.h. alle Aufgaben mit einer durchschnittlichen Durchführungszeit von 40-60 Minuten) durchzuführen, und folglich sei die Zeitersparnis durch eine geringere Anzahl zu kodierender Items nur geringfügig. Einschränkend hierzu lässt sich jedoch sagen, dass selbst eine reine Reduzierung der Kodierungsdauer (und nicht der Untersuchung) in der klinischen Praxis eine nicht zu vernachlässigende Zeitersparnis darstellen würde. Wenn man bspw. von einer Kodierungszeit von 30 Minuten für 31 Items ausgeht, so würde eine Kodierung von nur 5 Items die Kodierungszeit auf knapp 5 Minuten reduzieren, was eine Zeitersparnis von 25 Minuten pro Fall darstellen würde. Darüber hinaus haben jüngste Ergebnisse gezeigt, dass eine Kodierung der ADOS-Items bei Kindern auch in Bezug auf kürzere und unstrukturierte Verhaltensbeobachtungen von bis zu 10 Minuten Dauer erreicht werden konnte (Fusaro et al., 2014; Tariq et al., 2018). Diese Ergebnisse deuten darauf hin, dass auch eine Zeitersparnis im Zusammenhang mit der Untersuchung durchaus möglich sein könnte. Fusaro und Kolleg*innen (2014) haben beispielsweise die Anwendbarkeit der ADOS Modul 1 Item-Kodierungen auf kurze (<10 Minuten) unstrukturierte Heimvideos untersucht (d.h. es wurde nicht die gesamte ADOS-Untersuchung und Verhaltensbeobachtung durchgeführt). Insgesamt wurden hierzu 100 YouTube-Videos von Kindern (1-15 Jahre) mit selbstberichteten ASD- und non-ASD-

Diagnosen gesichtet und kodiert. Die Autor*innen berichten eine sehr hohe Klassifizierungsgenauigkeit von knapp 97%, mit 94% Sensitivität und 100% Spezifität. Obwohl nicht alle ADOS-Items auf die YouTube-Videos anwendbar waren, konnten die Autor*innen dennoch feststellen, dass ein Großteil der Items kodiert werden konnte. Insbesondere diejenigen Items aus Modul 1, die auch im Modul 4 enthalten sind und die den von uns identifizierten Merkmalen entsprechen (d.h. ungewöhnlicher Blickkontakt (B1), auf andere gerichtete Mimik (B2), Gestik (A9) und Qualität der sozialen Annäherungen (B9)), waren nach Angaben der Autor*innen leicht zu erkennen (die Häufigkeit der „nicht zutreffend“ Bewertungen bei diesen Items lag bei <0.044% pro Video). Diese Ergebnisse deuten darauf hin, dass selbst kurze und unstrukturierte Interaktionen ausreichend Informationen zur Kodierung der ADOS-Items und folglich der Erkennung von ASD liefern könnten. Aufbauend auf diese Arbeit haben Tariq und Kolleg*innen (2018) kürzlich untersucht, wie die reduzierten Item-Modelle aus früheren ML-Experimenten zu ADOS Modulen 1 bis 3 (Kosmicki et al., 2015; Levy et al., 2017; Wall et al., 2012) in die klinische Praxis umgesetzt werden könnten. Zu diesem Zweck wurde ein mobiles Webportal eingerichtet und Video-Rater*innen wurden angeleitet, die zuvor identifizierten, verkürzten Item-Algorithmen in kurzen Heimvideos (<5 Minuten) von Kindern mit und ohne ASD zu kodieren. Die Ergebnisse zeigen, dass alle Video-Rater*innen eine mittlere Bewertungszeit von 4 Minuten benötigten, um die Items zu kodieren, und ASD mit einer hohen Genauigkeit erkennen konnten (AUC bei 90%). Obwohl diese Ergebnisse durch eine unausgewogene Stichprobengröße, eine Vergleichsgruppe von typisch entwickelten Kindern sowie durch eine ASD-Selbstdiagnose eingeschränkt sind, erweisen sie sich dennoch als vielversprechend und sind ein wichtiger erster Schritt zur Umsetzung von ML-basierten Verhaltensmodellen in die klinische Praxis. Darüber hinaus unterstützen diese Ergebnisse die Möglichkeit, ADOS Items bzw. Item-Subsets anhand kürzerer, unstrukturierter sozialer Interaktionen wie z.B. kurzen Heimvideos zu bewerten, was eine erhebliche Zeitersparnis im Vergleich zur gesamten ADOS-Untersuchung darstellen und auch neue Möglichkeiten für ein Erst-Screening eröffnen würden.

In Anbetracht dieser Ergebnisse könnten auch die in unserer Studie identifizierten reduzierten 5-Item-Algorithmen in kürzeren Formaten wie unstrukturierten Videos oder kürzeren Interaktionen mit Kliniker*innen Anwendung finden. Hierdurch ließe sich die damit verbundene Durchführungs- und Beurteilungszeit deutlich reduzieren. Weitere

Untersuchungen sind zweifellos erforderlich, um zu beurteilen, ob sich die in dieser Arbeit identifizierten Verhaltensmerkmale als wesentlich und stabil in ihrer Differenzierungsfähigkeit zwischen Jugendlichen und Erwachsenen mit ASD und anderen komplexen klinischen Störungsbildern erweisen, und ob sie auch in kürzeren sozialen Interaktionen mit Jugendlichen und Erwachsenen angemessen beurteilt werden können. Nichtsdestotrotz unterstützen unsere Befunde die Hypothese, dass eine präzise ASD-Erkennung bei Jugendlichen und Erwachsenen bereits anhand einer reduzierten Anzahl von Verhaltensmerkmalen durchgeführt werden kann, wodurch die Komplexität des diagnostischen Prozesses reduziert werden könnte.

Darüber hinaus könnten die identifizierten Merkmale die Grundlage für die Entwicklung neuer und ökonomischer diagnostischer Screening-Tools darstellen, die speziell auf diesen Kernmerkmalen aufbauen. In einer Folge-Studie zu der vorliegenden Arbeit ist geplant, die identifizierten Merkmale an komplett neuen Daten zu evaluieren und als Screening-Instrument zu validieren. Hierzu sollen die identifizierten Merkmale im Rahmen eines internet-basierten Training-Tools eingearbeitet werden und Anwender*innen (d.h. Erstüberweiser*innen wie Allgemeinmediziner*innen oder Psychotherapeut*innen) zur Verfügung gestellt werden. Dieses Training-Tool soll neben einer Schulung in der korrekten Anwendung des Screening-Instruments weitere Informationen und Hilfen zur Versorgung von Patient*innen mit ASD im deutschen Gesundheitssystem enthalten. Eine Verbesserung des Screening-Prozesses könnte dazu beitragen, die Auslastung der ASD-Spezialambulanzen zu reduzieren, in dem die Entscheidungsfindung, wer zu einer weiteren ausführlichen (kosten- und zeitaufwendigen) diagnostischen Untersuchung an eine Spezialambulanz überwiesen wird, weiter optimiert werden könnte. Hierdurch ließe sich insgesamt die Wartezeit Betroffener auf Diagnostik-Termine verringern und in Folge auch eine frühzeitigere Planung passender und benötigter psychosozialer Interventionen realisieren.

Stärken und Grenzen

Die Hauptstärke dieser Studie ist die große und ausgewogene klinische Inanspruchnahmestichprobe von Jugendlichen und Erwachsenen mit ASD und Personen mit anderen klinischen Störungsbildern. Bei allen Personen ohne ASD-Diagnose lag initial ein Verdacht auf ASD vor, der nach einer ausführlichen diagnostischen Untersuchung ausgeschlossen werden konnte. Die non-ASD Gruppe

umfasste somit eine diagnostisch sehr vielfältige und anspruchsvolle Stichprobe mit ASD-ähnlichen Verhaltensmerkmalen, die jedoch durch das Vorliegen anderer psychiatrischer Grunderkrankungen wie affektiven Störungen, Angststörungen, ADHS und/oder Persönlichkeitsstörungen oder keiner aktuellen psychiatrischen Diagnose erklärt werden. Darüber hinaus bestand unsere Stichprobe sowohl aus männlichen als auch aus weiblichen Proband*innen und wies eine weite Altersspanne auf.

Bei der Auswertung der Alters-Subgruppen konnten wir leichte Unterschiede in den ausgewählten Items finden. Es ist möglich, dass sich dieser Effekt auch auf andere Subgruppen übertragen lässt. Lai und Kolleg*innen (2017) konnten beispielsweise zeigen, dass sich geschlechtsspezifische Unterschiede in der Symptombildung von Männern und Frauen mit ASD finden lassen. Zukünftige Studien sollten demnach auch mögliche Unterschiede in spezifischen Geschlechtsgruppen sowie spezifischeren klinischen Vergleichsgruppen (z.B. Persönlichkeitsstörungen, Angststörungen, andere Entwicklungsstörungen) untersuchen.

Darüber hinaus bestand unsere Stichprobe ausschließlich aus hochfunktionalen Personen, bei denen die ASD-Diagnose erst spät gestellt wurde. Es lässt sich vermuten, dass diese Personen in Bezug auf ihre Symptomausprägung eher zum milderen Ende des Spektrums gehören. Daher können unsere Ergebnisse nicht auf das gesamte Autismus-Spektrum verallgemeinert werden, insbesondere nicht auf Personen mit einer geringeren intellektuellen Funktionsfähigkeit.

Kritisch ist zudem zu bewerten, dass die Aussagekraft unserer Ergebnisse möglicherweise dadurch eingeschränkt wird, dass unsere Zielvariable (*best estimate* klinische Diagnose von ASD vs. non-ASD) nicht vollständig unabhängig ist von den Merkmalen, die für die Erstellung der Modelle verwendet wurden, da die ADOS-Items mit in die klinische Diagnosefindung einbezogen wurden. Dieses „Zirkularitäts-Problem“ wurde bereits in früheren Arbeiten diskutiert (für eine ausführlichere Diskussion siehe Langmann et al., 2017 und Kosmicki et al., 2015). Jedoch gibt es derzeit keine Möglichkeiten, dieses Problem zufriedenstellend zu lösen, da kein unabhängiges externes Kriterium für ASD (z.B. biologische Marker) verfügbar ist. Es gilt jedoch einschränkend zu betonen, dass, obwohl die ADOS in der Regel in die klinische Entscheidungsfindung einbezogen wurde, nicht alleinig hierdurch die Diagnose bestimmt wurde.

Schlussfolgerung und Ausblick

Zusammenfassend konnten in der vorliegenden Arbeit reduzierte Subsets von Verhaltensmerkmalen aus ADOS Modul 4 für Jugendliche und Erwachsene identifiziert werden, die vergleichbare Klassifizierungsleistungen aufwiesen wie bestehende Algorithmen, dabei aber eine signifikante Reduzierung der zu kodierenden Items darstellen. Obwohl alle Items der ADOS relevante Verhaltenskonzepte erfassen, umfassen die identifizierten Items möglicherweise wesentliche Konstrukte, die besonders gut zwischen Personen mit ASD und Personen mit anderen klinischen Störungsbildern differenzieren. Künftige Studien sollten untersuchen, ob sich diese Ergebnisse auch auf völlig neue Daten verallgemeinern lassen. Zukünftig könnten diese Ergebnisse dazu beitragen, den komplizierten (differential-)diagnostischen Prozess von ASD bei Jugendlichen und Erwachsenen zu vereinfachen, indem bestehende Algorithmen überarbeitet und verkürzt werden. Auch könnten neue diagnostische Screening-Instrumente auf Grundlage dieser wesentlichen Merkmale entwickelt werden.

9. Literaturverzeichnis

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th edition*. Arlington, VA: American Psychiatric Publishing.

Breiman, L. (2001). Random forests. *Machine learning*. **45**(1), 5-32. doi: doi.org/10.1023/A:1010933404324

Bone, D., Goodwin, M.S., Black, M.P., Lee, C.C., Audhkhasi, K., & Narayanan, S. (2015). Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises. *Journal of Autism and Developmental Disorders*. **45**, 1121–1136. doi: <https://doi.org/10.1007/s10803-014-2268-6>

Bone, D., Bishop, S.L., Black, M.P., Goodwin, M.S., Lord, C., & Narayanan, S.S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. *Journal of Child Psychology and Psychiatry*. **57**(8), 927-937. doi: 10.1111/jcpp.12559

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. **3**, 3, 223-230. doi: 10.1016/j.bpsc.2017.11.007

Carpenter, J. & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in medicine*. **19**(9), 1141-1164. doi: 10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F

DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. **44**(3), 837-845.

Fusaro, V.A., Daniels, J., Duda, M., DeLuca, T.F., D'Angelo, O., Tamburello, J., Maniscalco, J., & Wall, D.P. (2014). The potential of accelerating early detection of autism through content analysis of YouTube videos. *PLoS ONE*. **9**(4), e93533. doi: 10.1371/journal.pone.0093533

Happé, F.G., Mansour, H., Barrett, P., Brown, T., Abbot, P., & Charlton, R.A. (2016). Demographic and cognitive profile of individuals seeking a diagnosis of autism spectrum disorder in adulthood. *Journal of Autism Developmental Disorders*. **46**, 3469–3480. doi: 10.1007/s10803-016-2886-2

Hayes, J., Ford, T., Rafeeqe, H. & Russell, G. (2018). Clinical practice guidelines for diagnosis of autism spectrum disorder in adults and children in the UK: a narrative review. *BMC Psychiatry*. **18**(1), 222. doi: 10.1186/s12888-018-1800-1

Howlin, P., Moss, P., Savag, S. & Rutter, M. (2013). Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *Journal of the American Academy of Child & Adolescent Psychiatry*. **52**(6), 572–581. doi: 10.1016/j.jaac.2013.02.017

Hus, V. & Lord, C. (2014). The Autism Diagnostic Observation Schedule, Module 4: Revised algorithm and standardized severity scores. *Journal of Autism and Developmental Disorders*. **44**, 1996–2012. doi: 10.1007/s10803-014-2080-3

Joshi, G., Wozniak, J., Petty, C., Martelon, M.K., Fried, R., Bolfek, A., Kotte, A., Stevens, J., Furtak, S.L., Burgeois, M., Caruso, J., Caron, A., & Biederman, J. (2013). Psychiatric comorbidity and functioning in a clinically referred population of adults with autism spectrum disorders: A comparative study. *Journal of Autism and Developmental Disorders*. **43**(6), 1314–1325. doi: 10.1007/s10803-012-1679-5

Kamp-Becker, I., Albertowski, K., Becker, J., Ghahreman, M., Langmann, A., Mingebach, T., Poustka, L., Weber, L., Schmidt, H., Smidt, J., Stehr, T., Roessner, V., Kucharczyk, K., Wolff, N., & Stroth, S. (2018). Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *European Child & Adolescent Psychiatry*. **27**, 1193-1207. doi: 10.1007/s00787-018-1143-y

Kamp-Becker, I., Poustka, L., Bachmann, C., Ehrlich, S., Hoffmann, F., Kanske, P., Kirsch, P., Krach, S., Paulus, F.M., Rietschel, M., Roepke, S., Roessner, V., Schad-Hansjosten, T., Singer, T., Stroth, S., Witt, S., & Wermter, A.-K. (2017). Study protocol

of the ASD-Net, the German research consortium for the study of autism spectrum disorder across the lifespan: from a better etiological understanding, through valid diagnosis, to more effective health care. *BMC Psychiatry*. **17**(1), 206. doi: 10.1186/s12888-017-1362-7

Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. (2004). kernlab-an S4 package for kernel methods in R. *Journal of statistical software*. **11**(9), 1-20. doi: 10.18637/jss.v011.i09

Kosmicki, J., Sochat, V., Duda, M. & Wall, D. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational Psychiatry*. **5**(2), 514. doi: 10.1038/tp.2015.7

Kuhn, M. (2008). Caret package. *Journal of statistical software* **28**(5), 1-26. doi: 10.18637/jss.v028.i05

Küpper, C., Stroth, S., Wolff, N., Hauck, F., Kliwer, N., Schad-Hansjosten, T., Kamp-Becker, I., Poustka, L., Roessner, V., Schultebrasucks, K. & Roepke, S. (2020). Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. *Scientific Reports* **10**, 4805. doi: <https://doi.org/10.1038/s41598-020-61607-w>

Lombardo, M.V., Barnes, J.L., Wheelwright, S.J. & Baron-Cohen, S. (2007). Self-referential cognition and empathy in autism. *PLoS ONE*. **2**(9), e883. doi: 10.1371/journal.pone.0000883

Lai, M.C., Lombardo, M.V. & Baron-Cohen, S. (2014). Autism. *Lancet*. **383**, 896–910. doi: 10.1016/S0140-6736(13)61539-1

Lai, M.C. & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *Lancet Psychiatry*. **2**(11), 1013-27. doi: 10.1016/S2215-0366(15)00277-1

Lai, M.C., Lombardo, M., Ruigrok, A.N.V., Chakrabarti, B., Auyeung, B., Szatmari, P., Happé, F., Baron-Cohen, S., & MRC AIMS Consortium (2017). Quantifying and exploring camouflaging in men and women with autism. *Autism*. **21**(6), 690-702. doi: 10.1177/1362361316671012

Langmann, A., Becker, J., Poustka, L., Becker, K. & Kamp-Becker, I. (2017). Diagnostic utility of the autism diagnostic observation schedule in a clinical sample of adolescents and adults. *Research in Autism Spectrum Disorders*. **34**, 34–43. doi: 10.1016/j.rasd.2016.11.012

Le Couteur, A., Haden, G., Hammal, D. & McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: the ADI-R and the ADOS. *Journal of Autism and Developmental Disorders*. **38**(2), 362–372. doi: 10.1007/s10803-007-0403-3

Levy, S., Duda, M., Haber, N. & Wall, D.P. (2017). Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Molecular Autism*. **8**, 65. doi: 10.1186/s13229-017-0180-6

Lord, C., Risi, S., Lambrecht, L., Cook, E.H. jr, Leventhal, B.L., DiLavore, P.C., Pcikles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule–Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*. **30**, 205–223. doi: 10.1023/A:1005592401947

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part 1) Modules 1-4*. Torrance, CA: Western Psychological Services.

Maddox, B.B., Brodtkin, E.S., Calkins, M.E., Shea, K., Mullan, K., Hostager, J., Mandell, D.S., & Miller, J.S. (2017). The accuracy of the ADOS-2 in identifying autism among adults with complex psychiatric conditions. *Journal of Autism and Developmental Disorders*. **47**(9), 2703-2709. doi: 10.1007/s10803-017-3188-z.

Pepe, M., Longton, G. & Janes, H. (2009). Estimation and comparison of receiver operating characteristic curves. *The Stata Journal*. **9**(1), 1. doi: 10.1177/1536867X0900900101

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. **12**, 77. doi: 10.1186/1471-2105-12-77

Rutter, M., Le Couteur, A. & Lord, C. (2003). *Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles: Western Psychological Services.

Thabtah, F. (2018). Machine Learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health and Social Care*. **44**, 3, 278-297. doi: 10.1080/17538157.2017.1399132

Thabtah, F. & Peebles, D. (2019). A new machine learning model based on induction of rules for autism detection. *Health Informatics Journal*. doi: 10.1177/1460458218824711

Tariq, Q., Daniels, J., Schwartz, J.N., Washington, P., Kalantarian, H., & Wall, D.P. (2018). Mobile detection of autism through machine learning on home video: a development and prospective validation study. *PLOS Medicine*. **15**(11), e1002705. doi: 10.1371/journal.pmed.1002705

Tromans, S., Chester, V., Kiani, R., Alexander, R. & Brugha, T. (2018). The Prevalence of autism spectrum disorders in adult psychiatric inpatients: A systematic review. *Clinical Practice & Epidemiology in Mental Health*. **14**, 177-187. doi: 10.2174/1745017901814010177

Wall, D., Kosmicki, J., Deluca, T., Harstad, E. & Fusaro, V. (2012). Use of machine learning to shorten observation-based screening and diagnosis of autism. *Translational Psychiatry*. **2**(4), 100. doi: 10.1038/tp.2012.10

Wall, D.P., Dally, R., Luyster, R., Jung, J.Y & DeLuca, T.F. (2012). Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PloS ONE*. **7**(8), 43855. doi: 10.1371/journal.pone.0043855

Wigham, S., Rodgers, J., Berney, T., Le Couteur, A., Ingham, B., & Parr, J.R. (2018). Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: A systematic review. *Autism*. **23**, 287-305. doi: 10.1177/1362361317748245

World Health Organization (2016). *The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)*. Geneva: World Health Organization.

Zander, E., Willfors, C., Berggren, S., Choque-Olsson, N., Coco, C., Elmund, A., Moretti, A.H., Holm, A., Jifält, I., Kosieradzki, R., Linder, J., Nordin, V., Olafsdottir, K., Poltrago, L., & Bölte, S. (2016). The objectivity of the Autism Diagnostic Observation Schedule (ADOS) in naturalistic clinical settings. *European child & adolescent psychiatry*. **25**(7), 769-780. doi: 10.1007/s00787-015-0793-2

10. Eidesstattliche Versicherung

„Ich, Charlotte Küpper, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema „Autismus-Spektrum-Störungen im Jugend- und Erwachsenenalter: Kann die Diagnostik mit Hilfe von Verfahren des Maschinellen Lernens weiter optimiert werden?“ (englisch „Autism-spectrum-disorders in adolescence and adulthood: is it possible to further optimize the diagnostic process using machine learning?“) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe. Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet. Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht. Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem Erstbetreuer, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte. Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Charlotte Küpper

11. Anteilserklärung

Publikation: Küpper, C., Stroth, S., Wolff, N., Hauck, F., Kliewer, N., Schad-Hansjosten, T., Kamp-Becker, I., Poustka, L., Roessner, V., Schultebrasucks, K. & Roepke, S. (2020). Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. *Scientific Reports* 10, 4805. doi: <https://doi.org/10.1038/s41598-020-61607-w>

Beitrag im Einzelnen: Die von Charlotte Küpper in Erst-Autorenschaft erstellte Publikation entstand im Rahmen einer BMBF-geförderten Multi-Center-Studie unter Beteiligung von vier Studienzentren im Forschungsverbund „ASD-Net“. Für die oben genannte Publikation hat Charlotte Küpper folgenden Beitrag geleistet:

Ethikantrag: Die Verfassung und Einreichung des Ethikantrages wurde von Charlotte Küpper unter Supervision von Prof. Dr. Stefan Röpke durchgeführt.

Planung und Organisation: Charlotte Küpper war gemeinsam mit Dr. Sanna Stroth und Dr. Nicole Wolff zuständig für die Organisation, Koordination und Protokollierung regelmäßiger Telefonkonferenzen zum Austausch der vier Studienzentren zum Stand der Dateneingabe, -prüfung, -verarbeitung, -auswertung und -veröffentlichung.

Literaturrecherche: Die Literaturrecherche wurde von Charlotte Küpper durchgeführt.

Datenerhebung und Vorverarbeitung: Die Datenerhebung erfolgte zwischen 2001 und 2015 im Rahmen der Routine-Diagnostik an den vier beteiligten Studienzentren (spezialisierte Autismus-Ambulanzen in Deutschland). Charlotte Küpper führte während ihrer Mitarbeit an der Autismus-Ambulanz der Charité – Universitätsmedizin, Campus Benjamin Franklin, selbständig die Autismus-spezifische Diagnostik mittels Diagnostischer Beobachtungsskala für Autistische Störungen (ADOS), Diagnostischem Interview für Autismus – Revidiert (ADI-R) und klinischem Interview nach DSM-Kriterien durch. Zudem führte sie die differentialdiagnostische Untersuchung mittels Strukturiertem Klinischen Interview (SKID-I und -II), Fragebögenverfahren sowie Intelligenzdiagnostik mittels Wortschatztest (WST) durch.

Pro Zentrum erfolgte die Dateneingabe in eine Datenbank und es erfolgte eine anschließende Zusammenführung der Datenbanken in einen gemeinsamen Gesamt-

Datensatz. Anschließend erfolgte eine Plausibilitätsprüfung und eine Aufteilung des Datensatzes nach ADOS-Modul (1 bis 4). Charlotte Küpper war hauptverantwortlich für die Plausibilitäts-Prüfung und Vorverarbeitung der Modul 4 Daten. In diesem Rahmen führte sie die Umkodierung der Items und die Erstellung von altersspezifischen Sub-Datensätzen durch.

Datenanalyse: Charlotte Küpper hat die Datenanalyse in Zusammenarbeit mit Dr. Katharina Schultebraucks durchgeführt. Dr. Katharina Schultebraucks war vorwiegend für die Umsetzung der Machine Learning Verfahren zuständig und hat nach Rücksprache mit Charlotte Küpper die Abbildungen 1-3 erstellt. Charlotte Küpper hat die weiteren statistischen Analysen (Erfassung möglicher Gruppen-Unterschiede) vorgenommen und die Tabellen 1-5 erstellt. Charlotte Küpper hat in Zusammenarbeit mit Dr. Katharina Schultebraucks und Prof. Dr. Stefan Röpke alle Ergebnisse der statistischen Analysen interpretiert und kritisch hinterfragt, Übersichten der Daten erstellt und die Ergebnisse in den Gesamtzusammenhang eingeordnet.

Erstellen des Manuskripts: Charlotte Küpper war hauptverantwortlich für die Verfassung und Einreichung des Artikels, den sie unter Supervision von Prof. Dr. Stefan Röpke und Dr. Katharina Schultebraucks erstellte.

Unterschrift, Datum und Stempel des erstbetreuenden Hochschullehrers
(Prof. Dr. med. Stefan Röpke)

Unterschrift der Doktorandin
(Charlotte Küpper)

12. Auszug aus der Journal Summary List (ISI Web of KnowledgeSM)

Im Fachbereich „Multidisciplinary Sciences“ ist das Journal *Scientific Reports* an 15. Stelle von 69 nach Impact Factor sortierten Journals gelistet (entspricht 21.74%). Damit ist es unter den oberen 25 Prozent angesiedelt. Das Journal verfügt über einen Impact Factor von 4.01. Mit einem Eigenfaktor von 1.061540 gehört es zu den Topjournals.

Im Folgenden findet sich ein Screenshot eines Auszugs aus der „Journal Summary List“ (ISI Web of KnowledgeSM) im Fachbereich „Multidisciplinary Sciences“, dargestellt sind die ersten 18 von 69 Journals, sortiert nach Impact Factor (Stand 2018):

Journal Data Filtered By: **Selected JCR Year: 2018** Selected Editions: SCIE,SSCI
 Selected Categories: **“MULTIDISCIPLINARY SCIENCES”** Selected Category
 Scheme: WoS

Gesamtanzahl: 69 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE	745,692	43.070	1.285010
2	SCIENCE	680,994	41.037	1.070190
3	National Science Review	1,842	13.222	0.006500
4	Science Advances	21,901	12.804	0.110010
5	Nature Communications	243,793	11.878	1.103290
6	Nature Human Behaviour	1,230	10.575	0.006550
7	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	661,118	9.580	1.022190
8	Science Bulletin	3,569	6.277	0.009840
9	Scientific Data	3,240	5.929	0.015610
10	Frontiers in Bioengineering and Biotechnology	1,994	5.122	0.006540
11	Journal of Advanced Research	2,691	5.045	0.004780
12	Research Synthesis Methods	1,932	5.043	0.005420
13	GigaScience	2,674	4.688	0.012510
14	Annals of the New York Academy of Sciences	46,385	4.295	0.025840
15	Scientific Reports	302,086	4.011	1.061540
16	Journal of the Royal Society Interface	12,933	3.224	0.029190
17	NPJ Microgravity	203	3.111	0.000670
18	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES	19,227	3.093	0.028200

13. Druckexemplar der ausgewählten Publikation: Küpper, C., Stroth, S., Wolff, N., Hauck, F., Kliewer, N., Schad-Hansjosten, T., Kamp-Becker, I., Poustka, L., Roessner, V., Schultebrasucks, K. & Roepke, S. (2020). Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. *Scientific Reports* 10, 4805.

Küpper, C., Stroth, S., Wolff, N., Hauck, F., Kliewer, N., Schad-Hansjosten, T., Kamp-Becker, I., Poustka, L., Roessner, V., Schultebrasucks, K.* & Roepke, S.* (2020).

*shared senior authorship

Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning.

Scientific Reports, 10, 4805. doi: <https://doi.org/10.1038/s41598-020-61607-w>

OPEN

Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning

Charlotte Küpper^{1*}, Sanna Stroth², Nicole Wolff³, Florian Hauck⁴, Natalia Kliewer⁴, Tanja Schad-Hansjosten⁵, Inge Kamp-Becker², Luise Poustka⁶, Veit Roessner³, Katharina Schultebrucks^{7,8,9} & Stefan Roepke^{1,9*}

Diagnosing autism spectrum disorders (ASD) is a complicated, time-consuming process which is particularly challenging in older individuals. One of the most widely used behavioral diagnostic tools is the Autism Diagnostic Observation Schedule (ADOS). Previous work using machine learning techniques suggested that ASD detection in children can be achieved with substantially fewer items than the original ADOS. Here, we expand on this work with a specific focus on adolescents and adults as assessed with the ADOS Module 4. We used a machine learning algorithm (support vector machine) to examine whether ASD detection can be improved by identifying a subset of behavioral features from the ADOS Module 4 in a routine clinical sample of $N = 673$ high-functioning adolescents and adults with ASD ($n = 385$) and individuals with suspected ASD but other best-estimate or no psychiatric diagnoses ($n = 288$). We identified reduced subsets of 5 behavioral features for the whole sample as well as age subgroups (adolescents vs. adults) that showed good specificity and sensitivity and reached performance close to that of the existing ADOS algorithm and the full ADOS, with no significant differences in overall performance. These results may help to improve the complicated diagnostic process of ASD by encouraging future efforts to develop novel diagnostic instruments for ASD detection based on the identified constructs as well as aiding clinicians in the difficult question of differential diagnosis.

Autism Spectrum Disorders (ASD) comprise a range of pervasive neurodevelopmental disorders with a population prevalence of approximately 1%¹. They are characterized by early-onset persistent impairments in social communication and interaction as well as the presence of restricted, repetitive behaviors or interests^{2,3}. Diagnosing ASD is a complicated, lengthy and time-consuming process, which requires outstanding and specific clinical expertise^{4,5}. Although research makes constant progress in understanding the underlying genetic and neurobiological factors associated with ASD, there are currently no reliable biological markers for ASD and the diagnosis remains based on behavioral symptoms^{1,6,7}. The current so-called “gold standard” of ASD diagnosis comprises the use of various standardized diagnostic instruments that assist clinicians in reaching a best-estimate clinical diagnosis^{7–9}. Two of the most widely used diagnostic instruments are the Autism Diagnostic Observation Schedule (ADOS respectively ADOS-2 for the revised second edition)^{10,11} and the Autism Diagnostic Interview –

¹Department of Psychiatry, Charité - Universitätsmedizin Berlin, Campus Benjamin Franklin, Berlin, Germany.

²Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Philipps University, Marburg, Germany. ³Department of Child and Adolescent Psychiatry, TU Dresden, Dresden, Germany. ⁴Department of Information Systems, Freie Universität Berlin, Berlin, Germany. ⁵Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim/University of Heidelberg, Mannheim, Germany. ⁶Department of Child and Adolescent Psychiatry, University Medical Center, Göttingen, Germany. ⁷Department of Psychiatry, New York University School of Medicine, New York, USA. ⁸Vagelos School of Physicians and Surgeons, Department of Emergency Medicine, Columbia University Irving Medical Center, New York, USA. ⁹These authors jointly supervised this work: Katharina Schultebrucks and Stefan Roepke. *email: charlotte.kuepper@charite.de; stefan.roepke@charite.de

Revised (ADI-R)¹². The ADI-R is a semi-structured interview administered to parents or caregivers that focuses on developmental history and current symptom presentation. The ADOS is a standardized semi-structured diagnostic observation scale designed to assess important social-communicative behaviors as well as stereotypic and repetitive behavioral features. The ADOS includes four different modules for different age and language levels, with Module 4 intended for verbally fluent adolescents and adults. For each module, there is a diagnostic algorithm that allows for the classification of ASD or non-ASD^{10,11,13,14}. The ADOS is one of the psychometrically best-evaluated diagnostic tools in ASD, especially in children (Modules 1 and 2)^{4,15–17}, with fewer studies investigating the ADOS in adolescents and adults (Modules 3 and 4)^{18–21}. Although a good psychometric quality has been widely demonstrated, existing studies suggest a lower diagnostic utility in naturalistic clinical settings as well as with older individuals^{8,22–24}. While most cases with ASD are diagnosed in childhood, the diagnosis oftentimes remains unnoticed until adolescence and adulthood, particularly in those individuals with at least average cognitive and language abilities, better adaptive functions and more subtle symptom presentations^{25,26}. Diagnosing ASD in high-functioning adolescents and adults (i.e. without intellectual disability) can be even more challenging compared to childhood due to various factors: a care-giver based developmental history as acquired with the ADI-R is oftentimes unavailable in older individuals²⁵ and the individual's self-report of symptoms may be diminished due to impaired self-referential cognition²⁷. Additionally, learned compensatory skills might conceal impairments, thereby reducing diagnostic accuracy of observational tools such as the ADOS²⁵. Furthermore, most adults with ASD have at least one comorbid psychiatric disorder^{1,28} and ASD symptoms frequently overlap with those of other psychiatric conditions^{19,23} thereby further complicating differential diagnosis. Nonetheless, establishing an accurate and timely diagnosis is of great importance for those affected and the planning of suitable psychosocial interventions in order to promote positive outcome^{29–31}.

Previous studies have applied machine learning techniques to examine whether the process of diagnosing ASD can be improved by statistically identifying reduced subsets of features from existing diagnostic instruments reaching from self-administered screening questionnaires to clinician-administered diagnostic tools (for a recent overview, see Thabtah³²). A few authors have shown that efficiency and accessibility of existing pre-diagnostic screening questionnaires such as the Autism-Spectrum Quotient (AQ)^{33–35} or the Social Responsiveness Scale (SRS)^{36,37} can be improved using machine learning. Similar machine learning experiments have been run to identify subsets of behavioral features from clinician-administered diagnostic tools, namely ADOS (Module 1 to 3)^{38–42} and ADI-R^{36,39,43}. Findings of these studies suggest that ASD detection in children can be achieved with substantially fewer items compared to the original ADOS and ADI-R algorithms while retaining high diagnostic accuracy, sensitivity and specificity. To our knowledge, this has not yet been examined in a sample of adolescents and adults as assessed with the ADOS Module 4. The purpose of the present study therefore was to expand on the existing literature with a specific focus on high-functioning adolescents and adults. Advances of the present study are a large and balanced routine clinical sample of adolescents and adults with best-estimate clinical diagnoses of ASD and relevant psychiatric differential diagnoses, thereby being a good representation of the actual population presenting to ASD assessment settings. We aim to identify the diagnostically most informative features from the ADOS Module 4 that accurately differentiate between individuals with ASD and individuals with other clinically complex presentations using a data-driven machine learning approach. Although all items of the ADOS focus on relevant behavioral concepts, some items may be more discriminative and have higher classification ability particularly in a challenging clinical sample of older individuals that were all initially suspected of ASD. Identifying essential subsets of behavioral features that distinguish ASD from non-ASD cases could contribute to an enhancement of the complex diagnostic process in multiple ways: to improve existing diagnostic tools (i.e. revise existing classification algorithms), to shorten existing diagnostic tools such as the ADOS (for a critical discussion, however, see Bone and colleagues³⁶) and/or to inform the development of novel diagnostic tools and methods for initial screening based on these essential constructs.

Materials and Methods

Data sample and preprocessing. The study was conducted as a part of the ASD-Net, a research network with focus on ASD that is funded by the German federal ministry of education and research⁴⁴. All participant data came from four specialized ASD outpatient clinics in Germany where current diagnostic gold standard procedures had been applied to confirm or rule out a diagnosis of ASD. All participants were referred by specialists or self-referred to the outpatient departments. Participant data was collected retrospectively from the medical records of the respective clinic (retrospective chart review) and combined into one dataset for analysis. This procedure was approved by the Charité – Universitätsmedizin Berlin ethics committee (EA4/129/19) and due to the retrospective nature of data collection and analysis based on routinely obtained clinical data, the need for informed consent was waived by the local ethics committee. All methods were performed in accordance with the relevant institutional and international research guidelines and regulations.

The diagnostic procedure involved a standardized behavior observation in all cases (ADOS Module 4¹⁰), a standardized interview if parental informants were available (ADI-R¹²; care-givers were available in 62% of all cases (ASD: 71%, non-ASD: 50%)) and a differential diagnostic examination (established Structured Questionnaires and Structural Clinical Interviews frequently used in German-speaking countries), which aided trained and experienced clinicians in reaching a best-estimate clinical diagnosis. Multiple assessments were available for some of the cases, however, only the most recent assessment was considered for each case.

Our sample included data from 673 cases, of whom 57% received a diagnosis of ASD (“ASD”, $n = 385$) and 43% did not receive a diagnosis of ASD but relevant differential diagnoses such as affective disorders, anxiety disorders, ADHD and/or personality disorders or no current psychiatric diagnosis (“non-ASD”, $n = 288$; for a more detailed description of the phenotypic diversity see Supplementary Table 1). ASD subtypes according to ICD-10 (F84.0, F84.1, F84.5) were grouped together, giving us a binary outcome measure of “ASD” and “non-ASD” classes

Characteristic	ASD (n = 385)	non-ASD (n = 288)	Statistical test
Mean Age (SD)	25.63 years (11.27)	26.81 years (12.45)	n.s. ($t(582.92) = 1.27, p = 0.21$)
% age ≥ 18 years [n]	67% [n = 258]	68% [n = 196]	n.s. ($\chi^2(1) = 0.082, p = 0.78$)
% age > 21 years [n]	52.7% [n = 203]	51.7% [n = 149]	n.s. ($\chi^2(1) = 0.065, p = 0.80$)
Gender: % male [n]	74.3% male [n = 286]	72.9% male [n = 210]	n.s. ($\chi^2(1) = 0.16, p = 0.69$)
Mean IQ (SD)*	104.68 (16.00) (based on n = 343)	104.84 (15.49) (based on n = 245)	n.s. ($t(586) = 0.12, p = 0.90$)

Table 1. Sample Description. Abbreviation: n.s., non significant; ASD, autism spectrum disorder; SD, standard deviation. *Complete IQ data were available for 87% of the entire sample.

Code	Feature Description	ADOS core domain
A4*	Stereotyped/Idiosyncratic Use of Words or Phrases	Language/Communication
A8*	Conversation	Language/Communication
A9*	Descriptive, Conventional, Instrumental, or Informational Gestures	Language/Communication
A10	Emphatic or Emotional Gestures	Language/Communication
B1*	Unusual Eye Contact	Reciprocal Social Interaction
B2*	Facial Expressions Directed to Others	Reciprocal Social Interaction
B6	Empathy/Comments on Others' Emotions	Reciprocal Social Interaction
B8	Responsibility	Reciprocal Social Interaction
B9*	Quality of Social Overtures	Reciprocal Social Interaction
B10*	Quality of Social Response	Reciprocal Social Interaction
B11*	Amount of Reciprocal Social Communication	Reciprocal Social Interaction

Table 2. The 11 features from the ADOS Module 4 algorithm and the 5 features identified by the feature selection process for the whole sample (bold). Abbreviation: ADOS, Autism Diagnostic Observation Scale. *Items that are also comprised in the 12-item subset identified by Kosmicki and colleagues⁴¹. Further items that were identified by Kosmicki *et al.* that are not comprised in the ADOS algorithm are A7 (reporting of events), D1 (unusual sensory interest in play material/person), D2 (hand and finger and other complex mannerisms) and D4 (excessive interest in unusual or highly specific topics or objects).

for our machine learning procedures. There was no significant difference between the two groups regarding age, gender and IQ (Table 1).

The ADOS is a standardized observation scale designed to capture important social-communicative behaviors and stereotypic and repetitive behavioral features¹⁰. In Module 4, which is intended for verbally fluent adolescents and adults, these aspects are coded on 31 different items. Codes fall on an ordinal scale from 0 (no abnormality related to autism) to 2 (definite evidence of abnormality) and sometimes 3 (profound severity), with additional codes of 7 and 8 for abnormal behavior or behavior not exhibited during the observation, and a code of 9 for missing values (i.e. answers omitted or left blank).

The ADOS Module 4 provides a scoring algorithm consisting of a subset of the diagnostically most informative 11 items (see Table 2) from the Social Interaction and Communication domains for calculating a comparison score, which yields an instrument classification of autism, autism-spectrum or non-spectrum.

For data preprocessing for our machine learning analyses, we recoded ADOS codes of 3 to 2, and codes of 7 and 8 to 0 analogue ADOS handbook. Missing values (i.e. codes of 9) were imputed using k nearest neighbor imputation with $k = 5$ (*knnImpute*) using the build-in *preprocess()* function from the caret R package⁴⁵. In our dataset, six items were missing in 4–10% of all cases (items A6, B4, C1, E1, E2, E3), with all other items missing answers in less than 2.5% of the cases (for a more detailed description on the distribution of ADOS codes and missing values in our sample, see Supplementary Table 2). Furthermore, all numerical variables were normalized to range [0;1].

Machine learning. Previous classification experiments have utilized various machine learning techniques including support vector machines, tree-based models and general linear models³². In these previous publications, support vector machines (SVM) were among the models that performed best^{36,41,42}. Furthermore, SVM is one of the most frequently used algorithms that has been utilized for ASD classification due to its high predictive power³². Therefore, we decided to use SVM classification with radial kernel using *svmRadial* of the caret R package⁴⁵ as our machine learning classifier. We performed an additional analysis using random forest, which showed slightly lower predictive performance. Due to readability and space constraints, we only present results for SVM. Results from our random forest analysis can be found in the supplement (see Supplementary Table 3).

All 31 ADOS items were used as features and the individuals' best-estimate clinical diagnosis was used as our prediction class (ASD vs. non-ASD). All steps of data inspection and preprocessing, including imputation and analysis, were performed using R version 3.5.1 in Rstudio 1.1.456.

	Classification task	Training set	Test set	Total
All ages (n = 673)				
Positive events	ASD	n = 289	n = 96	N = 385
Negative events	non-ASD	n = 216	n = 72	N = 288
Adolescents \leq 21 years (n = 321)				
Positive events	ASD	n = 137	n = 45	N = 182
Negative events	non-ASD	n = 105	n = 34	N = 139
Adults > 21 years (n = 352)				
Positive events	ASD	n = 153	n = 50	N = 203
Negative events	non-ASD	n = 112	n = 37	N = 149

Table 3. Depiction of the classification task with the observed positive and negative events for the outcome in training and test set for the whole sample (“all ages”) as well as the age subgroups (“adolescents”, “adults”). Abbreviation: ASD, autism spectrum disorder.

To maximize the likelihood of unbiased results, rigorous guards against overfitting were implemented. The total sample was randomly split into a 75% partition as training set for building the models and hyperparameter testing, and a 25% partition as test set for evaluation of the predictive power of the final models in completely unseen cases not used to build the model (Table 3). The random split was stratified for the outcome variable to prevent data shift between training and test set. Potential information leakage was mitigated during the missing values imputation method, i.e. information about variable distributions secretly spreading from the training set data to the test set data was avoided by performing all preprocessing steps multiple times. At each cross-validation step, preprocessing (value range normalization and *knn*-imputation) was performed separately and anew for the resampled training set folds during the model building process and once for the test set before application of the final models for prediction. The dependent variable was removed from the dataset prior to imputation to prevent information leakage. Thus, it was made sure that for imputation at each fold during cross-validation, the information about the distribution of any predictor variable in the training set was unaffected by the distribution of the same variable in the test set and of the outcome variable.

Our machine learning approach consisted of multiple steps: First, we conducted a feature selection on the training set to identify a reduced feature subset with similar predictive performance profiles. We used recursive feature selection (*rfFuncs*) via the caret R package⁴⁵ applying random forests⁴⁶ to identify the most important features. The metric for hyperparameter tuning was Cohen’s kappa coefficient (equally weighing sensitivity and specificity), and 5-times repeated 10-fold cross-validation was applied. The second step consisted of training our reduced feature model on the training set. During model training, 5-times repeated 10-fold cross-validation was applied to mitigate the risk of overfitting and to achieve stable prediction results. The metric for hyperparameter tuning was the area under the ROC curve (AUC) used to select the model with the largest AUC on the training set. The set of tuned hyperparameters was *sigma* and *cost* with a random search of 500 different combinations. All other tunable parameters were set to default values⁴⁷. The third step was testing our reduced feature model on the dedicated test set to see how well the reduced model can separate ASD from non-ASD in completely unseen data of a separate hold-out set, i.e. data that was never used to build the model.

To compare model performances, we evaluated the AUCs of the predictions attained from our newly identified reduced feature subset vs. all 31 items of the ADOS vs. the subset of 11 items proposed by the ADOS algorithm. Additionally, we compared model performance to the 12-item classifier previously identified by Kosmicki and colleagues⁴¹ in their experiments looking at children and adolescents as assessed with ADOS Module 3 (for a list of these items, see Table 2). For evaluation of the differences in terms of AUC, we relied on DeLong’s test⁴⁸ for two correlated ROC curves as well as a bootstrap resampling significance test^{49,50} for two correlated ROC curves (each time comparing the overlap of the confidence intervals with 10 000 bootstrapped iterations).

Because of an overall large age distribution in our sample (ages ranging from 10 to 72 years, with a median age of 22 years), all of the above-mentioned steps were performed in the whole sample (“all ages”, N = 673) as well as in age subgroups of adolescents aged \leq 21 years (“adolescents”, n = 321, 56.7% ASD (n = 182)) and adults aged > 21 years (“adults”, n = 352, 57.7% ASD (n = 203)). For further information on the age distribution of our sample and the age subgroups, see Supplementary Table 4 and Supplementary Figures 1 to 3.

Results

Looking at the whole sample, our recursive feature selection algorithm selected five features as the most important ones, i.e. those which had on average the highest ability to predict adolescents and adults with ASD compared to adolescents and adults with other clinically complex presentations during cross-validation: Features A9 (Descriptive, Conventional, Instrumental or Informational Gestures), B1 (Unusual Eye Contact), B2 (Facial Expressions directed to Others), B10 (Quality of Social Response) and B11 (Amount of Reciprocal Social Communication). All of the selected five features correspond to the domains of Social Interaction and Communication of the ADOS and are comprised in the original 11-feature ADOS algorithm as well as in the 12-item classifier proposed by Kosmicki and colleagues⁴¹ (Table 2). Items A9, B1 and B2 depict abnormalities in the participant’s reciprocal nonverbal communication and interaction observed during the ADOS examination, whereas items B10 and B11 constitute qualitative summary items, wherein the clinician rates abnormalities in the participant’s overall social behaviors during the ADOS examination.

	SVM models			
	5-feature model*	11-feature model (ADOS algorithm)	31-feature model (all ADOS items)	12-feature model (Kosmicki <i>et al.</i> ⁴¹)
All ages (n = 673)				
Training Set				
AUC (Sensitivity, Specificity)	0.87 (0.72, 0.87)	0.87 (0.75, 0.88)	0.87 (0.73, 0.88)	0.87 (0.73, 0.85)
Test Set				
AUC (Sensitivity, Specificity)	0.82 (0.71, 0.83)	0.84 (0.85, 0.76)	0.84 (0.79, 0.81)	0.84 (0.77, 0.82)

Table 4. Performance of the machine learning models on the training and test set for the whole sample (“all ages”). Abbreviation: AUC, Area under the ROC curve; SVM, support vector machine. *5-feature model for “all ages”: A9, B1, B2, B10, B11.

With this reduced feature subset of only 5 features, we observed an AUC of 0.87 (sensitivity = 0.72, specificity = 0.87) in the trainings set, which is comparable to the performance of the 11-feature model of the ADOS algorithm, the 31-feature model using all items of the ADOS and the 12-item subset identified by Kosmicki and colleagues⁴¹ (all AUCs of 0.87; see Table 4).

For independent validation of our classifiers, we computed the performance of the models on the held-out test set. When independently predicting the best-estimate clinical diagnosis, our models achieved slightly lower AUCs (see Table 4): our reduced 5-feature model achieved an AUC of 0.82 (sensitivity = 0.71, specificity = 0.83) compared to AUCs of 0.84 of the 11-feature model (sensitivity = 0.85, specificity = 0.76), the 12-feature model proposed by Kosmicki *et al.*⁴¹ (sensitivity = 0.77, specificity = 0.82) and the 31-feature model (sensitivity = 0.79, specificity = 0.81; Table 4, Fig. 1).

In a final step, we compared the models’ performances. No significant differences were found when comparing the AUCs of the reduced 5-feature model with the 11-feature model suggested by the ADOS algorithm (DeLong’s test: $Z = -1.63$, $p = 0.10$; bootstrapping: $D = -1.61$, $p = 0.11$, number of bootstrapped resampled = 10 000), the 12-feature model proposed by Kosmicki *et al.*⁴¹ (DeLong’s test: $Z = -1.27$, $p = 0.20$; bootstrapping: $D = -1.26$, $p = 0.21$, number of bootstrapped resampled = 10 000) and the 31-feature model (DeLong’s test: $Z = -1.29$, $p = 0.20$; bootstrapping: $D = -1.26$, $p = 0.21$, number of bootstrapped resamples = 10 000).

Results separately for age subgroups (“adolescents” and “adults”) can be found in Table 5. Compared to our whole sample feature subset (“all ages”: items A9, B1, B2, B10, B11), our recursive feature selection algorithm selected slightly different features as the most important ones for the specific age subgroups: items A9, B1, B2, B3, B9 for “adolescents” vs. items A9, B2, B3, B9, B10 for “adults”. Similar to the “all ages” feature subset, however, all of the selected features correspond to the domains of Social Interaction and Communication of the ADOS. Two items (B3 (Language Production and Linked Nonverbal Communication) and B9 (Quality of Social Overtures)) were included in both age subgroups but not in the “all ages” sample. Item B9 is a qualitative summary item that assesses the overall quality of the individual’s attempts to initiate social interactions, whereas item B3 holds information about how the individual’s vocalizations are accompanied by nonverbal behaviors (such as eye contact, gestures and facial expression).

By splitting our “all ages” sample into age subgroups, we were able to increase the overall prediction accuracy of our models (see Table 5). Comparable to the whole sample, there were no significant differences between the AUCs of the 5-feature models (“adolescents”: AUC = 0.90; “adults”: AUC = 0.84), the 11-feature model proposed by the ADOS algorithm (“adolescents”: AUC = 0.88; “adults”: AUC = 0.87), the 31-feature model (“adolescents”: AUC = 0.87; “adults”: AUC = 0.87) and the 12-feature model proposed by Kosmicki *et al.*⁴¹ (“adolescents”: AUC = 0.84; “adults”: AUC = 0.85) in the test sets in the respective subgroups. For a detailed depiction of the comparison tests’ results as well as ROC curves for the age subgroups see Supplementary Figures 4 and 5.

Discussion

Although the search for objective biological markers associated with ASD is currently under way, the current standard of ASD diagnosis remains based on behavioral symptoms alone. One of the most widely used behavioral examination tools which aids clinicians in reaching a best-estimate clinical diagnosis is the ADOS. Although effective, administration and evaluation of the ADOS is time-consuming and requires outstanding specific clinical expertise. However, accurate and timely identification of ASD in adulthood has become an important clinical issue^{25,31} and the field needs novel methods for initial screening as well as accurate diagnostic tools that can reach a larger percentage of cases⁸. Moreover, specificity of the ADOS in clinical settings is low, demonstrating the clinical difficulty of differentiating individuals with ASD from individuals with suspected ASD but other best-estimate psychiatric diagnoses or no psychiatric diagnoses. This is particularly challenging in high-functioning older individuals, due to symptom overlaps of various disorders with ASD, an increased comorbidity rate in ASD and the lack of reliable information about early development. Machine learning techniques have previously been applied to examine whether ASD detection can be improved by identifying core subsets of behavioral features that discriminate between children with and without ASD (ADOS Module 1 to 3). In the present study, we sought to expand on the existing literature with a specific focus on the differentiation of adolescents and adults with ASD from individuals with other clinically complex presentations (ADOS Module 4).

Using an SVM-based approach, we identified a reduced subset of 5 behavioral features from the ADOS Module 4 that showed good specificity (83%) and sensitivity (71%) on our whole sample. Furthermore, with an AUC of 82%, our reduced classifier reached performance close to that of the 11-feature algorithm proposed by the

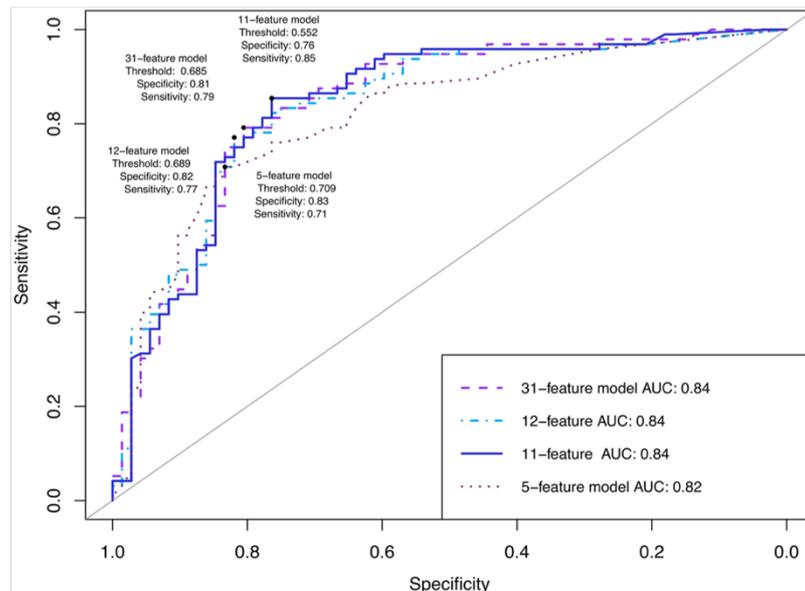


Figure 1. Receiver operating characteristic (ROC) curves evaluating the predictive power in the test set for the whole sample (“all ages”). Optimal ROC threshold with the highest sum of sensitivity + specificity is plotted⁶¹.

	SVM models			
	5-feature models*	11-feature model (ADOS algorithm)	31-feature model (all ADOS items)	12-feature model (Kosmicki <i>et al.</i> ⁴¹)
Adolescents ≤ 21 years (n = 321)				
Training Set				
AUC (Sensitivity, Specificity)	0.83 (0.67, 0.85)	0.85 (0.58, 0.92)	0.84 (0.66, 0.85)	0.85 (0.70, 0.86)
Test Set				
AUC (Sensitivity, Specificity)	0.90 (0.78, 0.88)	0.88 (0.87, 0.82)	0.87 (0.84, 0.79)	0.84 (0.84, 0.77)
Adults > 21 years (n = 352)				
Training Set				
AUC (Sensitivity, Specificity)	0.87 (0.69, 0.88)	0.88 (0.71, 0.89)	0.86 (0.65, 0.89)	0.86 (0.62, 0.92)
Test Set				
AUC (Sensitivity, Specificity)	0.84 (0.90, 0.76)	0.87 (0.92, 0.84)	0.87 (0.90, 0.84)	0.85 (0.90, 0.78)

Table 5. Performance of the machine learning models on the training and test set for the age subgroups “adolescents” (≤21 years) and “adults” (>21 years). Abbreviation: AUC, Area under the ROC curve; SVM, support vector machine. *5-feature model for “adolescents”: A9, B1, B2, B3, B9. *5-feature model for “adults”: A9, B2, B3, B9, B10.

ADOS manual and the full ADOS consisting of 31 features (both AUCs of 84%) with no significant differences in overall performance. Additionally, we evaluated model performance of the previously proposed 12-item-subset by Kosmicki *et al.*⁴¹ on our dataset and compared performance with our models. Results showed that our reduced 5-feature model and the proposed 12-feature model (AUC of 84%) achieved similar predictive performance, with no significant differences when comparing the AUCs. In their original publication, Kosmicki *et al.*⁴¹ applied SVM with a radial basis function and reported model performance of 97.7% sensitivity and 97.2% specificity on their whole sample (n = 1924 ASD, n = 214 non-ASD). For individuals for whom a best-estimate clinical diagnosis was available (n = 1568 ASD, n = 175 non-ASD), their 12-feature classifier displayed 99.1% sensitivity and 70.9% specificity. Independently testing this proposed 12-item-subset on our dataset, we found an overall performance of 77% sensitivity and 82% specificity, which is lower compared to their original findings. However, their proposed 12-feature subset still achieved an overall good prediction performance that is in line with performances of our reduced subset as well as the ADOS algorithm and the full ADOS. A possible explanation for the observed lower prediction performance might lie in the composition of the investigated samples. First of all, their classifier was built using ADOS Module 3 data, with Module 3 being intended for children and adolescents. Consequently, their sample was considerably younger compared to ours and displayed an overall lower developmental level. Second, there was a large imbalance in their data between groups (ASD vs. non-ASD), with our dataset being more balanced. Lastly, while our sample was a complex clinical and naturalistic sample of individuals seen for comprehensive diagnostic evaluation, data of their investigated sample came from archival repositories and

diagnosis was partly only based on ADOS results (thus missing a best-estimate clinical diagnosis). Nonetheless, our independent evaluation of their proposed classifier shows encouraging results and lends additional support to the hypothesis that models using minimal feature subsets from the ADOS are accurate for ASD classification.

By splitting up our sample further into age subgroups of adolescents (≤ 21 years) and adults (> 21 years), we were able to even further increase prediction accuracy of our abbreviated 5-feature subsets (“adolescents”: AUC of 90%; “adults”: AUC of 84%). Similar to the whole sample (“all ages”), we found no significant differences when comparing the overall prediction performance of the different models (“adolescents”: AUCs of 90% (5-feature) vs. 88% (11-feature) vs. 87% (31-feature) vs. 84% (12-feature proposed by Kosmicki *et al.*⁴¹); “adults”: AUCs of 84% (5-feature) vs. 87% (11-feature) vs. 87% (31-feature) vs. 85% (12-feature proposed by Kosmicki *et al.*⁴¹)) in the respective subgroups.

Having a closer look at the items selected by our feature selection algorithm, we find that all 5 items identified for the whole sample (“all ages”: A9, B1, B2, B10, B11) were part of the original 11-feature algorithm proposed by the ADOS manual, thereby supporting the relative diagnostic importance of these items and possibly preserving some of its diagnostic validity. Looking more closely at the age-subgroups, results show that the majority of items overlaps between the groups (“adolescents”: A9, B1, B2, B3, B9; “adults”: A9, B2, B3, B9, B10) and that all selected items stem from the ADOS’ domains of Social Interaction and Communication. Interestingly, there is a substantial overlap of 4 items between the two age subgroups, possibly reflecting their relative importance independent of developmental level. The outstanding difference between the two age subgroups were the inclusion of the more “basic” social-communicative item B1 (Unusual Eye Contact) in the adolescent group and the more “complex” qualitative and summative item B10 (Quality of Social Response) in the adult group. These findings suggest that adolescents with ASD show more saliently unusual eye contact compared to adults with ASD whereas adults show more “complex” social interaction impairments. This finding is in line with previous studies suggesting that there may be differences in ASD symptom presentation over the lifespan due to learnt compensatory behaviors and general developmental gains.^{13,51}

Comparing our findings with those of previously published minimal behavior subsets from ADOS Module 1 to 3, we find parallels in the features selected across different studies and age groups. Looking at the reduced 12-feature classifier proposed by Kosmicki and colleagues⁴¹ in their experiments on Module 3, we find a correspondence of all of the items identified in our classifiers apart from item B3. Regarding their experiments looking at ADOS Module 2⁴¹, there was an overlap of 4 items between our reduced classifiers and their identified 9-feature classifier (Gestures (A9), Unusual Eye Contact (B1), Quality of Social Overtures (B9) and Amount of Reciprocal Social Communication (B11)). Similar overlaps can be found looking at the experiments conducted by Levy and colleagues⁴², who identified reduced 5-feature classifiers for Module 2 and Module 3 with an overlap of 3 items for Module 2 (Unusual Eye Contact (B1), Facial Expressions Directed to Others (B2) and Quality of Social Response (B10)) as well as an overlap of 3 items for Module 3 (Facial Expressions Directed to Others (B2), Quality of Social Overtures (B9) and Quality of Social Response (B10)). Not yet published results of our own group also identified reduced 5-feature classifiers for Module 1, 2 and 3 using Support Vector Machine. Here, we find overlaps of feature B1 for Module 1 and features B2, B9 and B11 for Module 2 and 3 with our reduced classifier. Taken together, there are marked similarities in the features selected across different studies and age groups, particularly for items B1, B2, B9, B10 and B11, which all stem from the Social Interaction domain of the ADOS.

Impairments in social interaction in general are very common in many different psychiatric disorders. However, Bishop and colleagues⁵², for example, recently showed that the items stemming from the social interaction and communication domains of the ADOS Module 3 algorithm could be further separated into two subdomains: impairments in “Basic Social Communication” (such as eye contact, facial expression, gestures) and impairments in “Interaction Quality” including items that measure more complex aspects of social interaction (such as conversation, amount of reciprocal social communication, quality of rapport), with the former being impaired in children with ASD and mostly intact in children without ASD but other clinical diagnoses, and the latter being impaired in both groups. In line with this finding, Drimalla and colleagues⁵³ for example recently developed an alternative screening method for ASD detection and showed that it was possible to accurately detect ASD compared to healthy controls by automatically analyzing the subject’s facial expression, vocal features and gaze patterns during a video-based simulated social interaction. Although all items of the ADOS algorithm focus on relevant behavioral concepts of social-communication impairments, our abbreviated classifiers might contain relevant “core” features from the ADOS that are more specific to adolescents and adults with ASD. Thus, focusing on a reduced subset of items that are most specific to ASD and less influenced by other clinical presentations could assist clinicians in the difficult process of differential diagnosis.

Our findings suggest that reduced subsets of only 5 items are equally accurate in predicting ASD as the 11-item algorithm proposed by the ADOS manual or even the whole ADOS exam consisting of 31 items. Compared to the full ADOS, however, our abbreviated 5-feature classifiers present a substantial reduction of 84% of the number of behavioral features that have to be coded in order to assess ASD. It has been argued before³⁹, however, that administration time for the ADOS cannot simply be reduced by reducing the number of items coded, as ADOS items are not directly tied to any subtask. In order to code the items, it is always necessary to administer the whole-length ADOS exam (i.e. all subtasks) and consequently the time saved by coding only a smaller number of behavioral features is marginal. In contrast to this view, however, recent findings have shown that ASD classification of children can be achieved by applying the ADOS items to shorter, unstructured social interactions^{54–56} and even by relying solely on written extracts of children’s medical and educational records^{57,58}, thereby suggesting that a reduction in time associated with ASD detection might be feasible. For example, Fusaro and colleagues⁵⁴ have assessed the feasibility of applying all of the ADOS Module 1 codes, but not the behavior observation exam, to short (< 10 minutes) and unstructured home-videos collected from YouTube. They collected and rated 100 videos of children (age 1–15) with self-reported ASD and non-ASD diagnoses and achieved classification accuracy of about 97% (with 94% sensitivity and 100% specificity). While certainly not all ADOS items were relevant to the

YouTube videos, the authors did find, however, that a majority of the items could be applied. In particular, the items from Module 1 that overlap with the Module 4 items included in our identified 5-feature subsets (i.e. unusual eye contact (B1), facial expressions directed to others (B2), gestures (A9) and quality of social overtures (B9)) were readily detectable (frequency of “not applicable” ratings was at <0.044% per video). These findings suggest that short, unstructured interactions can provide sufficient information to rate ADOS codes and to detect ASD. Building upon this work, Tariq and colleagues⁵⁵ recently investigated how the reduced feature subsets from the ADOS modules 1 to 3 that were identified in previous machine learning experiments^{38,41,42} could be translated into clinical practice. For this purpose, they created a mobile web portal and asked video raters to assess the previously identified minimal feature subsets in short home videos (<5 minutes) of children with and without ASD. Results showed that all video raters took a median rating time of 4 minutes to detect ASD with high accuracy (AUC at 90%). Although constrained by an unbalanced sample size, a comparison group of typically-developing children as well as relying on self-report diagnosis of ASD, these results are nonetheless promising and an important first step in translating machine-learning based behavioral models into clinical practice. Furthermore, these results support the possibility of achieving ASD detection in briefer, unstructured social interactions such as home videos of under 5- to 10-minute length, which is considerably shorter than the full ADOS exam assessment time (which is on average 30–60 minutes).

Considering these findings, the minimal 5-feature subsets identified in our study might have the potential for utility in shorter format approaches such as assessment in unstructured home videos or brief interactions with clinicians, thereby possibly reducing redundancy in the existing diagnostic algorithm and accompanying assessment time. Furthermore, the features identified could inform the development of novel diagnostic screening tools that specifically build upon these core features. Further examination is certainly needed to evaluate whether the behaviors identified with our classifiers prove essential and stable in their ability to discriminate between adolescents and adults with ASD and other complex clinical presentations and whether they may also be adequately assessed in shorter social interactions with adolescents and adults. Nonetheless, these findings support the claim that accurate ASD detection in adolescents and adults can be performed using smaller sets of behavioral features, thereby potentially allowing for a reduction in the complexity of the diagnostic procedure.

Strengths and limitations. The main strength of this study is the large and balanced routine clinical sample of individuals all initially suspected of having ASD. The non-ASD group was a diagnostically very diverse and heterogeneous group, with ASD-like behavioral characteristics that originate, however, from different underlying psychiatric conditions such as affective disorders, anxiety disorders, ADHD and/or personality disorders or no current psychiatric diagnosis. Furthermore, there was a large age range in our sample and our sample consisted of male as well as female participants, thereby being a good representation of the routine clinical population presenting for ASD diagnostics in adolescence and adulthood. Evaluating age-restricted subgroups, we found slight differences in the items selected, hence the predictive value of our reduced algorithms might also differ in clinical as well as gender subgroups^{59,60}. Future studies should investigate differences in more specific clinical comparison groups (e.g. personality disorders, anxiety disorders, other developmental disorders) as well as specific gender groups.

Additionally, our sample consisted of high-functioning individuals of whom most presented late in life for an ASD diagnosis, thereby probably belonging to the mild end of the spectrum. Therefore, results cannot be generalized to the entire ASD spectrum, especially to those individuals with lower intellectual functioning.

Lastly, our outcome criterion (best-estimate clinical diagnosis of ASD vs. non-ASD) was not independent of the features utilized for building the prediction algorithm, therefore possibly confounding our results. This circularity problem has been previously discussed, however, there are currently no ways of satisfyingly addressing this issue as there is no independent external criterion available (for a more detailed discussion see^{20,41}). However, even though the ADOS was usually factored into the clinical decision making, it did not solely determine the diagnosis.

Conclusion

Taken together, our results are an important step forward towards improving ASD detection in older individuals and shed some light particularly in the difficult issue of differential diagnosis among clinically complex cases. We identified reduced subsets of behavioral features from the ADOS Module 4 for the whole sample as well as adolescents and adults separately that showed comparable classification performance to that of the full ADOS and the existing ADOS algorithm. While all items of the ADOS focus on relevant behavioral concepts, the items identified may have a higher ability to differentiate individuals with ASD from individuals with other clinically complex presentations in adolescence and adulthood. Although further studies are needed to evaluate these reduced classifiers' ability to generalize to completely new and unseen data and to determine its clinical value, these results may help to improve the complicated ASD diagnostic process in adolescents and adults by encouraging future efforts to improve existing diagnostic instruments such as the ADOS, thereby aiding clinicians particularly in the difficult question of differential diagnosis, as well as to develop novel diagnostic instruments for ASD detection.

Data availability

The datasets generated and analysed during the current study are not publicly available due to medical confidentiality but are available from the first author on reasonable request pending the approval of the coauthors.

Received: 17 October 2019; Accepted: 27 February 2020;

Published online: 18 March 2020

References

- Lai, M. C., Lombardo, M. V. & Baron-Cohen, S. Autism. *Lancet*. **383**, 896–910 (2014).
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 5th edn.* (American Psychiatric Publishing, Arlington, VA, 2013).
- World Health Organization. *The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)*. (World Health Organization, Geneva, 2016).
- Kamp-Becker, I. *et al.* Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *European Child & Adolescent Psychiatry*. **27**, 1193–1207 (2018).
- Charman, T. & Gotham, K. Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders – lessons from research and practice. *Child Adolesc Ment Health*. **18**(1), 52–63 (2013).
- Whyatt, C. P. & Torres, E. B. Autism Research: An objective quantitative review of progress and focus between 1994 and 2015. *Front. Psychol.* **9**, 1526 (2018).
- Hayes, J., Ford, T., Rafeeqe, H. & Russell, G. Clinical practice guidelines for diagnosis of autism spectrum disorder in adults and children in the UK: a narrative review. *BMC Psychiatry*. **18**(1), 222 (2018).
- Wigham, S. *et al.* Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: A systematic review. *Autism*. **23**, 287–305 (2018).
- Falkmer, T., Anderson, K., Falkmer, M. & Horlin, C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. *Eur Child Adolesc Psychiatry*. **22**, 329–40 (2013).
- Lord, C. *et al.* The Autism Diagnostic Observation Schedule–Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*. **30**, 205–223 (2000).
- Lord, C. *et al.* Autism Diagnostic Observation Schedule, Second Edition (ADOS–2) Manual (Part 1) Modules 1–4. Torrance, CA: Western Psychological Services (2012).
- Rutter, M., Le Couteur, A. & Lord, C. *Autism Diagnostic Interview-Revised (ADI-R)*. (Western Psychological Services, Los Angeles, 2003).
- Hus, V. & Lord, C. The Autism Diagnostic Observation Schedule, Module 4: Revised algorithm and standardized severity scores. *Journal of Autism and Developmental Disorders*. **44**, 1996–2012 (2014).
- Gotham, K., Risi, S., Pickles, A. & Lord, C. The Autism diagnostic observation schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*. **37**(4), 613–627 (2007).
- Le Couteur, A., Haden, G., Hammal, D. & McConachie, H. Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: the ADI-R and the ADOS. *J Autism Dev Disord*. **38**(2), 362–372 (2008).
- Mazefsky, C. A. & Oswald, D. P. The discriminative ability and diagnostic utility of the ADOS-G, ADI-R, and GARS for children in a clinical setting. *Autism*. **10**(6), 533–549 (2006).
- De Bildt, A. *et al.* Interrelationship between autism diagnostic observation schedule-generic (ADOS-G), autism diagnostic interview-revised (ADI-R), and the diagnostic and statistical manual of mental disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders*. **34**(2), 129–137 (2004).
- Bastiaansen, J. A. *et al.* Diagnosing autism spectrum disorders in adults: The use of Autism Diagnostic Observation Schedule (ADOS) Module 4. *Journal of Autism and Developmental Disorders*. **41**, 1256–1266 (2011).
- De Bildt, A., Sytema, S., Meffert, H. & Bastiaansen, J. A. C. J. The Autism Diagnostic Observation Schedule, Module 4: Application of the revised algorithms in an independent, well-defined, Dutch sample (n = 93). *Journal of Autism and Developmental Disorders*. **46**(1), 21–30 (2015).
- Langmann, A., Becker, J., Poustka, L., Becker, K. & Kamp-Becker, I. Diagnostic utility of the autism diagnostic observation schedule in a clinical sample of adolescents and adults. *Res Autism Spectr Disord*. **34**, 34–43 (2017).
- Fusar-Poli, L. *et al.* Diagnosing ASD in Adults Without ID: Accuracy of the ADOS-2 and the ADI-R. *J Autism Dev Disord*. **47**, 3370–3379 (2017).
- Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T. & Manning-Courtney, P. Use of the autism diagnostic observation schedule (ADOS) in a clinical setting. *Autism*. **15**(2), 143–162 (2013).
- Maddox, B. B. *et al.* The accuracy of the ADOS-2 in identifying autism among adults with complex psychiatric conditions. *J Autism Dev Disord*. **47**(9), 2703–2709 (2017).
- Zander, E. *et al.* The objectivity of the Autism Diagnostic Observation Schedule (ADOS) in naturalistic clinical settings. *European child & adolescent psychiatry*. **25**(7), 769–780 (2016).
- Lai, M. C. & Baron-Cohen, S. Identifying the lost generation of adults with autism spectrum conditions. *Lancet Psychiatry*. **2**(11), 1013–27 (2015).
- Happé, F. G. *et al.* Demographic and cognitive profile of individuals seeking a diagnosis of autism spectrum disorder in adulthood. *J Autism Dev Disord*. **46**, 3469–3480 (2016).
- Lombardo, M. V., Barnes, J. L., Wheelwright, S. J. & Baron-Cohen, S. Self-referential cognition and empathy in autism. *PLoS ONE*. **2**(9), e883 (2007).
- Joshi, G. *et al.* Psychiatric comorbidity and functioning in a clinically referred population of adults with autism spectrum disorders: A comparative study. *Journal of Autism & Developmental Disorders*. **43**(6), 1314–1325 (2013).
- Tromans, S., Chester, V., Kiani, R., Alexander, R. & Brugha, T. The Prevalence of autism spectrum disorders in adult psychiatric inpatients: A systematic review. *Clinical Practice & Epidemiology in Mental Health*. **14**, 177–187 (2018).
- Howlin, P. & Moss, P. Adults with autism spectrum disorders. *CanJPsychiatry*. **57**(5), 275–283 (2012).
- National Institute for Health and Care Excellence (NICE). Clinical guideline [CG142]. Autism spectrum disorder in adults: diagnosis and management. Published date: June 2012, last updated: August 2016., <https://www.nice.org.uk/guidance/cg142>.
- Thabtah, F. Machine Learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health and Social Care*. **44**(3), 278–297 (2018).
- Thabtah, F. & Peebles, D. A new machine learning model based on induction of rules for autism detection. *Health Informatics Journal*. (2019).
- Akter, T. *et al.* Machine learning-based models for early stage detection of autism spectrum disorders. *IEEE Access*. (2019).
- Omar, K. S., Mondal, P., Khan, N. S., Rizvi, M. R. & Islam, M. N. A machine learning approach to predict autism spectrum disorder. *International Conference on Electrical, Computer and Communication Engineering (ECCE)*. (2019).
- Bone, D. *et al.* Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. *J Child Psychol Psychiatry*. **57**(8), 927–937 (2016).
- Duda, M., Ma, R., Haber, N. & Wall, D. P. Use of machine learning for behavioral distinction of autism and ADHD. *Translational Psychiatry*. **6**, e732 (2016).
- Wall, D., Kosmicki, J., Deluca, T., Harstad, E. & Fusaro, V. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl Psychiatry*. **2**(4), 100 (2012).
- Bone, D. *et al.* Applying machine learning to facilitate autism diagnostics: Pitfalls and promises. *Journal of Autism and Developmental Disorders*. **45**, 1121–1136 (2015).
- Duda, M., Kosmicki, J. & Wall, D. Testing the accuracy of an observation-based classifier for rapid detection of autism risk. *Transl Psychiatry*. **4**(8), 424 (2014).

41. Kosmicki, J., Sochat, V., Duda, M. & Wall, D. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Transl Psychiatry*. **5**(2), 514 (2015).
42. Levy, S., Duda, M., Haber, N. & Wall, D. P. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Molecular Autism*. **8**, 65 (2017).
43. Wall, D. P., Dally, R., Luyster, R., Jung, J. Y. & DeLuca, T. F. Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PLoS ONE*. **7**(8), 43855 (2012).
44. Kamp-Becker, I. *et al.* Study protocol of the ASD-Net, the German research consortium for the study of autism spectrum disorder across the lifespan: from a better etiological understanding, through valid diagnosis, to more effective health care. *BMC Psychiatry*. **17**(1), 206 (2017).
45. Kuhn, M. Caret package. *Journal of statistical software* **28**(5), 1–26 (2008).
46. Breiman, L. Random forests. *Machine learning*. **45**(1), 5–32 (2001).
47. Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. kernlab—an S4 package for kernel methods in R. *Journal of statistical software*. **11**(9), 1–20 (2004).
48. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. **44**(3), 837–845 (1988).
49. Carpenter, J. & Bithell, J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in medicine*. **19**(9), 1141–1164 (2000).
50. Pepe, M., Longton, G. & Janes, H. Estimation and comparison of receiver operating characteristic curves. *The Stata Journal*. **9**(1), 1 (2009).
51. Howlin, P., Moss, P., Savag, S. & Rutter, M. Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *Journal of the American Academy of Child & Adolescent Psychiatry*. **52**(6), 572–581 (2013).
52. Bishop, S., Havdahl, K. A., Huerta, M. & Lord, C. Subdimensions of social-communication impairment in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*. **57**(8), 909–916 (2016).
53. Drimalla, *et al.* Detecting autism by analyzing a simulated social interaction. Conference: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. At: Dublin, Ireland (2019).
54. Fusaro, V. A. *et al.* The potential of accelerating early detection of autism through content analysis of YouTube videos. *PLoS ONE*. **9**(4), e93533 (2014).
55. Tariq, Q. *et al.* Mobile detection of autism through machine learning on home video: a development and prospective validation study. *PLOS Medicine*. **15**(11), e1002705 (2018).
56. Abbas, H., Garberson, F., Glover, E. & Wall, D. Machine learning approach for early detection of autism by combining questionnaire and home video screening. *Journal of the American Medical Association*. **25**(8), 1000–1007 (2018).
57. Maenner, M. J., Yeargin-Allsopp, M., Braun, K. V., Christensen, D. L. & Schieve, L. A. Development of a machine learning algorithm for the surveillance of autism spectrum disorder. *PLoS ONE*. **11**(12), e0168224 (2016).
58. Lee, S. H., Maenner, M. J. & Heilig, C. M. A comparison of machine learning algorithms for the surveillance of autism spectrum disorder. *PLoS ONE*. **14**(9), e0222907 (2019).
59. Lai, M. C. *et al.* Quantifying and exploring camouflaging in men and women with autism. *Autism*. **21**(6), 690–702 (2017).
60. Lai, M. C., Lombardo, M. V., Auyeung, B., Chakrabarti, B. & Baron-Cohen, S. Sex/gender differences and autism: setting the scene for future research. *J Am Acad Child Adolesc Psychiatry*. **54**(1), 11–24 (2015).
61. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. **12**, 77 (2011).

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF, grant number: FKZ 01EE1409A). We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité – Universitätsmedizin Berlin. KS was supported by the German Research Foundation (DFG, grant number: SCHU 3259/1–1). The authors would like to thank all individuals who participated in this study as well as all clinicians who assisted in the collection of data that was included in this study. Furthermore, we would like to thank Gerti Gerber for her assistance in the conduct of this research.

Author contributions

Principle investigators and site leaders who contributed to the design of the study are I.K.B., L.P., V.R. and S.R. C.K., S.S., N.W. and T.S.H. collected the data. The planning, design, and review of the formal analysis of the data were done by K.S., N.K. and F.H. and K.S. performed the statistical analysis, all authors contributed to the interpretation of the data. C.K., K.S. and S.R. wrote the manuscript, all authors contributed to reviewing and editing the manuscript and approved the final version of the manuscript.

Competing interests

L.P. has received conference attendance support or speaking fees from Shire. She receives research funding by the BMBF, DFG and EU and royalties from Hogrefe, Kohlhammer and Schattauer. V.R. has received payment for consulting and writing activities from Lilly, Novartis, and Shire Pharmaceuticals, lecture honoraria from Lilly, Novartis, Shire Pharmaceuticals, and Medice Pharma, and support for research from Shire Pharmaceuticals and Novartis. He has carried out clinical trials in cooperation with the Novartis, Shire, Servier and Otsuka companies. The remaining authors declare that they have no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-61607-w>.

Correspondence and requests for materials should be addressed to C.K. or S.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

14. Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

15. Komplette Publikationsliste

Veröffentlichungen in Fachzeitschriften mit Peer-Review

Tebartz van Elst, L., Schaller, U.M., Fangmeier, T., Hennig, O., Kieser, M., Koelkebeck, K., **Küpper, C.**, Roessner, V., Wildgruber, D., & Dziobek, I. (*in preparation*). Evaluation of a three-arm, cluster-randomised controlled study to assess the effects of FASTER and SCOTT&EVA trainings for adults with high-functioning autism spectrum disorder (ASD).

Küpper, C., Stroth, S., Wolff, N., Hauck, F., Kliwer, N., Schad-Hansjosten, T., Kamp-Becker, I., Poustka, L., Roessner, V., Schultebrucks, K. & Roepke, S. (2020). Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. *Scientific Reports*, 10, 4805.

Bachmann, C., Höfer, J., Kamp-Becker, I., **Küpper, C.**, Poustka, L., Roepke, S., Roessner, V., Stroth, S., Wolff, N., & Hoffmann, F. (2019). Internalised stigma in adults with autism: A German multi-center survey. *Psychiatry Research*, 276, 94–99.

Höfer, J., Hoffmann, F., Kamp-Becker, I., **Küpper, C.**, Poustka, L., Roepke, S., Roessner, V., Stroth, S., Wolff, N., & Bachmann, C.J. (2019). Complementary and alternative medicine use in adults with autism spectrum disorder in Germany: results from a multi-center survey. *BMC Psychiatry*, 19(1):53.

Catarino, A., **Küpper, C.**, Werner-Seidler, A., Dalgleish, T., & Anderson, M. (2015). Failing to Forget: Inhibitory Control Deficits Compromise Memory Suppression in Post-Traumatic Stress Disorder. *Psychological Science*, 26(5), 604-616.

Küpper, C., Benoit, R.G., Dalgleish, T., & Anderson, M. (2014). Direct Suppression as a Mechanism for Controlling Unpleasant Memories in Daily Life. *Journal of Experimental Psychology: General*, 143(4), 1443-1449.

Gutz, L., **Küpper, C.**, Renneberg, B., & Niedeggen, M. (2011). Processing Social Participation: An Event-Related Brain Potential Study. *NeuroReport*, 22(9), 453-458.

16. Danksagung

Diese Dissertation ist das Ergebnis meiner mehrjährigen wissenschaftlichen Tätigkeit in der Autismus-Ambulanz der Charité - Universitätsmedizin Berlin, Campus Benjamin Franklin. Viele Menschen haben mich auf diesem Weg begleitet und dazu beigetragen, dass diese Promotionsschrift nun vorliegt. Im Folgenden möchte ich mich bei einigen von ihnen ganz herzlich bedanken.

Zunächst möchte ich mich bei der Klinik für Psychiatrie und Psychotherapie der Charité Universitätsmedizin Berlin, Campus Benjamin Franklin, und der leitenden Direktorin Frau Prof. Dr. Isabella Heuser für die Möglichkeit bedanken, diese Promotion durchzuführen. Ich möchte insbesondere Prof. Dr. Stefan Röpke danken, der meine Promotion mit seinem großen Fachwissen begleitet hat. Durch seine wertvollen Rückmeldungen und seinen kompetenten Rat konnte ich meine Arbeit ständig weiter entwickeln und verbessern. Ich konnte mich bei Bedarf auf seine fachliche Unterstützung verlassen, weiß aber auch wertzuschätzen, dass er mir wissenschaftliche Freiräume gewährte, so dass mir ein selbständiges Arbeiten möglich war. Ich möchte mich weiterhin bei Frau Dr. Katharina Schultebraucks bedanken, die mir mit ihrer Expertise in der statistischen Datenanalyse, insbesondere in Bezug auf die Verfahren des Maschinellen Lernens, und ihrer hilfsbereiten Art eine große Unterstützung war. Darüber hinaus möchte ich mich bei allen Co-Autor*innen und Mitarbeiter*innen der anderen Studienzentren bedanken. Ohne deren engagierten Einsatz, insbesondere von Dr. Sanna Stroth und Dr. Nicole Wolff, wäre diese Arbeit nicht denkbar gewesen. Entscheidend für die Durchführung der Studie war natürlich auch das Mitwirken aller Menschen mit Autismus-Spektrum-Störung, die sich bereit erklärt haben, an der Studie teilzunehmen - vielen Dank!

Danken möchte ich außerdem meinen Eltern, die mir kritisches Denken vorgelebt haben und mir durch ihre herzlichen Ermutigungen das Selbstvertrauen gegeben haben, dieses Projekt anzugehen. Zu guter Letzt möchte ich meinem Lebensgefährten danken, der mich immer bestärkt hat und der immer für mich da ist.