



LUNDS
UNIVERSITET

Institutionen för psykologi
Psykologprogrammet

Nio hinder för klinisk diagnostik av autismspektrumtillstånd med artificiell intelligens: en systematisk litteraturstudie

Max Kanakura
Psykologexamensuppsats, 2022

Handledare: Per Johnsson
Examinator: Petri Kajonius

Abstract

This systematic literature review has tried to answer how eye-tracking and adjacent technologies in combination with machine learning (n = 13 studies) or deep learning (n = 6 studies) respectively can be used for the purpose of autism spectrum diagnostics (ASD). Included articles have been published in peer-reviewed journals. Most studies have sample sizes below 100 participants (n = 14 studies), 101-160 participants (n = 3 studies), 161-1000 participants (n = 2 studies). Studies include toddlers (16 months) to adults. Machine learning tends to be less accurate (59-93%) than deep learning (81-95%) in sorting individuals with or without ASD. The highest accuracy, precision (positive predicative value) and specificity (reliability) in the included studies were achieved by comparing eye-scanning and EEG data (95%, 95% and 95% respectively). A list of 9 obstacles that scientific studies would need to address before these AI technologies could be implemented in clinical practice as a tool for identifying ASD are reviewed.

Keywords: Machine / deep learning, autism spectrum disorder, artificial intelligence, diagnostics.

Sammanfattning

Det här är en systematisk litteraturöversikt som har prövat att besvara hur ögonskanning och angränsande teknologier i kombination med maskin- (n = 13 studier) och djupinlärning (n = 6 studier), var för sig, kan användas för att diagnostisera autismspektrumstörning (AST). Inkluderade artiklar har varit publicerade i vetenskapligt granskade journaler. De flesta studierna har haft under 100 deltagare (n = 14 studier), 101-160 deltagare (n = 3 studier), 161-1000 deltagare (n = 2 studier). Studierna har inkluderat småbarn (16 månader) till vuxna personer. Maskininlärning tenderar att vara mindre träffsäkert (59-93%) än djupinlärning (81-95%) i att särskilja individer med och utan AST. Den högsta träffsäkerheten, precisionen (positivt prediktionsvärde) och specificiteten (tillförlitlighet) i en inkluderad studie uppnåddes med ögonskanning och EEG data (95%, 95% och 95%). En lista med 9 hinder som vetenskapliga studier skulle behöva övervinna innan AI teknologi kan införas inom klinisk praktik med syfte att diagnostisera patienter med AST har sammanställts.

Nyckelord: Maskininlärning, djupinlärning, autismspektrumstörning, artificiell intelligens, diagnostik.

Tack

Jag vill tacka min handledare Per Johnsson för handledning, stöd och uppmuntran under skrivandet av uppsatsen.

Innehållsförteckning

| | |
|--|----|
| Nio hinder för klinisk diagnostik av autismspektrumtillstånd med artificiell intelligens: en systematisk litteraturstudie | 1 |
| Forskningsrelevans | 1 |
| Teoretiska begrepp | 2 |
| Vad är ögonskanning och hur används det? | 2 |
| Vad är datorsyn? | 3 |
| Vad är artificiell intelligens, AI? | 4 |
| Vad är maskininlärning? | 4 |
| Vad är djupinlärning? | 6 |
| Artificiell intelligens (AI) och autismspektrumstörningar (AST) | 8 |
| Syfte | 10 |
| Frågeställningar | 10 |
| Metod | 11 |
| Urval | 11 |
| Bedömningskriterier för AI | 14 |
| Resultat | 16 |
| Studier som inriktar sig på kännetecken och använder maskininlärning | 16 |
| Studie som avser kännetecken och använder djupinlärning | 18 |
| Studier som särskiljer AST / inte AST med maskininlärning | 19 |
| Studier som urskiljer AST / inte AST med djupinlärning | 30 |
| Diskussion | 35 |
| Framtida forskning | 42 |
| Slutsatser | 43 |
| Referenser | 44 |

Nio hinder för klinisk diagnostik av autismspektrumtillstånd med artificiell intelligens: en systematisk litteraturstudie

Forskningsrelevans

Syftet med att starta en utredning för autismspektrumtillstånd (AST) är ofta en strävan för att förstå varför en persons beteende uppfattas som avvikande av omgivningen. Ordet AST består av tre delar, *autism*, som avser ett annorlunda fungerande i ett socialt samspel. *Spektrum*, som beskriver en kontinuerlig skala. Samt *tillstånd*, som påvisar att det handlar om ett varande och inte en sjukdom (1177 vårdguiden, u.å). Sammansättningen av de tre orden avser att påvisa att det finns stora variationer i hur AST tar sig uttryck.

Symtom på AST utvecklas ofta i tidig ålder, vilket gör att det ses som ett neurologiskt utvecklingstillstånd (Mazumdar et al., 2020). Symtom som mindre spontan uppmärksamhet för sociala stimuli kan ses redan vid 6 månaders ålder (Chang et al., 2021). Fler symtom börjar märkas i 2-4 års ålder (Alcañiz et al., 2021). Andra typiska tecken är försämrad verbal och icke verbal kommunikation, begränsade, repetitiva beteenden och påverkan på motoriska funktioner (Vabalas et al., 2020).

Väntetiden på en AST utredning kan vara lång. Idag krävs att sofistikerade och tidskrävande undersökningar genomförs på en klinik (Zhao et al., 2021). En utredning kan innebära flera återbesök, då det ofta ingår ett flera tester, undersökningar och observationer.

Beteendeskattningar utförda av kliniker anses idag vara det mest tillförlitliga sättet att ställa neurologiska diagnoser som AST (Minissi et al., 2021; Sapiro et al., 2019). Problematiskt är att AST bedömningar är subjektiva, kräver betydande klinisk expertis och är beroende av hur väl den naturliga miljön som hem och skola fungerar (Sapiro et al., 2019). Med subjektiva bedömningar finns risk för bias. Barn och anhöriga kan komma med svar som anses socialt gynnsamma i andras ögon beträffande symtom (Minissi et al., 2021). Det finns flera förhoppningar om vad diagnostisering med artificiell intelligens (AI) ska kunna tillföra, ett är att få bedömningskriterierna för AST att bli mer objektiva och träffsäkra (He et al., 2021; Li et al., 2020; Minissi et al., 2021; Sapiro et al., 2019). Befintliga utredningsmetoder för AST har svårt att screena större populationer (Minissi et al., 2021; Sapiro et al., 2019). En förväntan är att AI ska bidra till att metoder för screening och diagnostik av AST blir mer effektiva (He et al., 2021; Kojovic et al., 2021). Med diagnostisering i större skala kan sannolikt väntetider för AST utredning minskas (Vabalas et al., 2020). Ordet ”förväntningar” ger uttryck för en framtidsvision. Vilka hinder behöver

övervinnas för att AI teknik med ögonskanning ska övergå från en vision till en klinisk tillämpning? Det är frågan som det fokuseras på i den här studien.

Det första utkast som nästan färdigställas undersökte om hur artificiell intelligens kunde diagnostisera fyra psykiatriska diagnoser. Nämligen attention deficit hyperactivity disorder (ADHD), autism, bipolärsjukdom och depression. Urvalet av artiklar kunde inte försvaras som ett systematiskt inhämtade, därför kasserades utkast och processen började om. Kunskapen som inhämtades har gjort det här arbetet bättre.

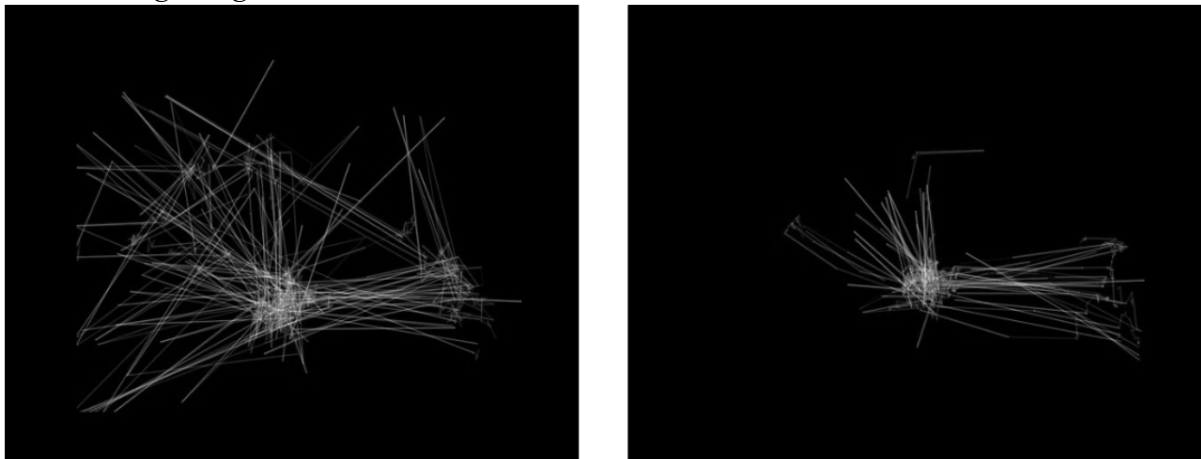
Teoretiska begrepp

Vad är ögonskanning och hur används det?

Ögonskanning sker med specialglasögon som med infraröd teknik kan registrera ögonrörelser.¹ Ordet ”ögonskanning” kommer fortsättningsvis att avse bärandet av den här typen av glasögon. Figur 1 visar en illustration från en studie av Cilia et al. (2021) som

Figur 1.

Visualisering av ögonrörelser.



Exempel på visualisering av spårade ögonrörelserna som Cilia et al. (2021) redovisar i sin studie. Till vänster en person med AST och till höger en person utan AST.

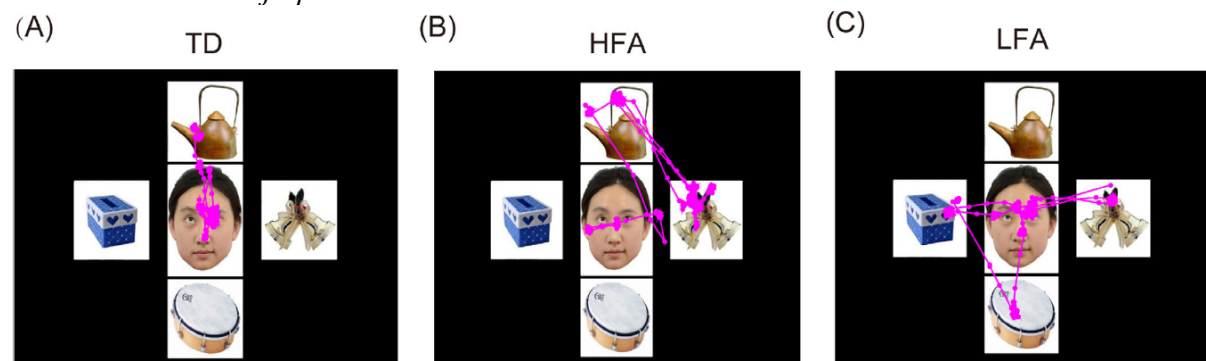
visualiserar spårning av ögonrörelser för en person med AST (vänster bild) och motsvarande för en person utan AST (höger bild). Det som framkommer av bilderna är att de två individerna har haft väldigt olika ögonrörelser när de har ombetts att genomföra samma typ av uppgift. Ögonskanning kan utöver att registrera hur ögonen rör sig, processa insamlade data för att avgöra vad det är som glasögonbäraren iakttar.

¹ Ett svenskt företag som bedriver egen forskning och utveckling av ögonspårningsglasögon är Tobii.

Fokus på iakttagande kan exemplifieras med en studie av He et al. (2021). De har grupperat barn i tre kategorier, normalutvecklade barn, barn med högfunderade autism och barn med lågfunderade autism. Barnen visas en uppsättning med 5 bilder, en central bild och en bild på var sida om den centrala bilden. Den centrala bilden avser att ge barnet vägledning om vart den ska rikta blicken. Om barnet lyckas får barnet en verbal uppmuntran. Om barnet misslyckas händer inget. Figur 2 visar resultatet från ett barn från vardera kategorin. Bilderna

Figur 2.

Rikta blicken med hjälp av ledtrådar.



Den centrala bilden i mitten avser ger en ledtråd om vart barnet ska rikta sin blick i He et al.:s (2021) studie. Linjerna i bilden visar hur barnets blick har rört sig under momentet när den illustrerade bilden visades. Till vänster visas resultatet för ett barn med normal utveckling (typical developed). I mitten visas ögonrörelser för ett barn för högfunderande autism (HFA) och till vänster ett barn med lågfunderade autism (LFA).

åskådliggör att barnen har haft väldigt skilda ögonrörelser och olika förmåga att läsa av och ta hjälp av ledtråden. Helheten i bilden är informativ på två sätt, (1) genom att det går att avläsa hur mycket fokus som ägnats åt att titta på ledtråden. Den informationen framkommer genom att iakta hur mycket barnet har tittat i området kring ögonen på den centrala bilden i mitten. (2) Har barnet följt vägledningen som ledtråden avsåg att ge? Om ja, i vilken grad? Svaret framträder när linjen följs som spårar ögonens rörelser mellan bilderna (He et al., 2021).

Vad är datorsyn?

En angränsande teknik är datorsyn (eng. computer vision), som använder vardaglig kameraoptik som sitter på vardagselektronik som en läsplatta, en smart mobiltelefon eller på datorns bildskärm. Tekniken möjliggör att insamlingen av data sker på avstånd förutsatt att det finns fri sikt mellan kameran och den som observeras.² Utöver ett fokus på ögonrörelser kan datorsyn användas för övervakning av en persons rörelsemönster och registrering av känslouttryck (Sapiro et al., 2019). När rörelsemönster registreras och undersöks i samband

² Idén med datorsyn (computer vision) är att få datorer att skilja mellan olika objekt utifrån hur de ser ut, precis som människor kan ta till sig information från en visuell värld (IBM, u.å).

med AST är det fokus på hur lederna koordineras (Kojovic et al., 2021), vilket kan kategoriseras som kinematik (Vabalas et al., 2020).

Tekniker som mäter hjärnans elektriska aktivitet vid hjärnskalpen (Elektroencefalografi, EEG [Zhang et al. 2021b]) eller hjärnavbildningstekniker som magnetresonanstomografi (MRI [Xu et al. 2021]) kräver besök på sjukhus (Li et al., 2020). Fördelen med ögonskanningstekniker är att diagnostisering av AST kan påbörjas i hemmiljö (Zhao et al., 2021). Argument för datorsyn, är att utrustning för ögonskanning kan vara dyr och kräver kalibrering av instrumentet (Chang et al., 2021). Zhao et al. (2021) har genomfört sin studie med bärbar teknik för ögonskanning, ändå menar de att datorsyn är att föredra då kalibreringsmomentet blir överflödigt.

Vad är artificiell intelligens, AI?

Det kan beskrivas som ett induktivt tänkande som identifierar regler för övergripande mönster i stora datamängder, så att ny kunskap genereras (Tai et al., 2019). En av AI:s styrkor är att det går snabbt att identifiera mönster, ett motsvarande manuellt arbete skulle vara tidskrävande (Çaliyurt, 2020). Bland forskare användes begreppet AI för att söka anslag, mellan kollegor används i stället termen ”maskininläring” (Crawford, 2021). En angränsande mindre kategori inom AI är djupinläring (Zhang et al., 2021a). Djupinläring har sina egna algoritmer för maskininläring och lär sig i högre grad att klassificera information direkt från rådata (Quaak et al., 2021). Båda metoderna kan användas för att finna kännetecken som särskiljer olika kategorier ifrån varandra. AI i kombination med tekniker för ögonskanning har visat sig tillämpbart för att urskilja individer med och utan AST från tidig ålder (Chang et al., 2021) till vuxna individer (Lio et al., 2020).

Vad är maskininläring?

Standardiserad maskininläring är en transparent modell där ingångsdata viktas och processas till en matematisk användbar funktion.³ Det som gör modellens funktion transparent, är att det i ekvationen kan utläsas vilken betydelse varje inmatat värde tillskrivs och slutligen används funktionen för att beräkna resultatet vid nya testningar (Quaak et al., 2021). Här kommer maskininläring användas i bemärkelsen *”en algoritm som ”sjävlär” hur den bäst kan återskapa ett dataset, med syfte att användas till kategorisering eller prediktion i framtiden”* (de Saint Laurent, 2018).⁴

³ Maskininläring av data har varit framgångsrikt inom medicinsk diagnostik för dermatologi (hudsjukdomar), radiologi (bilddiagnostik) och patologi (läran om sjukdomar [Çaliyurt, 2020])

⁴ Egen översättning.

Användning av maskininlärning är en process med sju delsteg, (1) datainsamling, (2) förberedande databearbetning, (3) visualisering av data och analys, (4) val av kännetecken och teknik, (5) träning av en tillämpningsmodell, (6) utvärdering av modellen och (7) prediktion (Kumar et al., 2021). För att säkerställa att det material som används för att träna artificiell intelligens håller kvalitativt, krävs enligt Washington et al. (2019) en mänsklig uppmärkning av råmaterial. Uppmärkning av material är den största flaskhalsen för den digitala utvecklingen. En lösning är en fördelning av arbetsbördan genom digitalt folkligt arbete för att märka upp råmaterial (Washington et al., 2019).⁵

Det finns en risk i att överlåta uppmärkning av material till allmänheten, då AI är känslig för två saker. Mängden träningsdata som processas och datasetets kvalitet, båda faktorerna har en avgörande betydelse för förmågan till god framtida prediktion. Om datainsamling sker på skilda platser eller registreras utifrån olika normer, kan det ha en direkt inverkan på resultatet. Problem med bias kan uppstå om träningsmaterialet är skevt eller alltför enahanda (de Saint Laurent, 2018). En sådan risk är om träningsurvalet till exempel endast innehåller pojkar med AST diagnos och en motsvarande grupp flickor saknas.

Enligt Graham et al. (2019) dominerar tre typer av maskininlärning inom vård och hälsa, det är övervakad, oövervakad och djupinlärning. Ett kännetecken för övervakad maskininlärning (Supervised Machine Learning, SML), är att inkluderad data har namngetts och grupperats enligt ett kategoriseringssystem, till exempel personer med AST och kontrollgrupp. Vid tillämpning ges en algoritm uppgiften att träna sig själv att identifiera kännetecken med de på förhand namngivna kategorier i en större samling av "träningsdata". I fasen med träningsdata fungerar de namngivna kategorierna som vägledning för algoritmens självinlärning.

I den påföljande fasen ska algoritmens förmåga att göra korrekta kategoriseringar värderas på ett nytt "testdata" material. Det är viktigt att träningsdata hålls oberoende från testdatamaterialet (Quaak et al., 2021). Om förmågan att korrekt skilja mellan önskade kategorier har försämrats i körningen av testdata, anses algoritmen ha blivit överanpassad till sin träningsdata, vilket indikerar en låg extern validitet. Ett lågt deltagarantal i förhållande till mängden inkluderade kännetecken ökar risken för felkategorisering (Vabalas et al., 2020). Om fler deltagare kan inkluderas i en studie minskar risken för överanpassning och instrumentets bedömningsförmåga stabiliseras (Minissi et al., 2021). En utmaning för studier

⁵ Washington et al., 2019 använder ordet "crowdsourcing" och hänvisar till webbsidan Amazon Mechanical Turk.

inom artificiell intelligens med inriktning på mental hälsa, är den begränsade tillgången till stora dataset (Lee et al., 2021). Det förekommer att forskare delar på befintliga dataset, men forskare som utgår från samma dataset gör gärna egna kvalitets- och preferensurval, vilket gör att inkluderad data kan komma att skilja åt. Det försvårar för direkta jämförelser av resultat (Quaak et al., 2021). Det finns algoritmer som påminner om befintliga statistiska metoder och andra som är unika för övervakad maskininläring (Graham et al., 2019).

Vid oövervakad maskininläring, (Unsupervised Machine Learning, UML), saknas förbestämda kategoriseringar. När algoritmen identifierar kännetecken eller strukturer i datamaterialet sorteras de till kluster. För att bedöma materialets användbart behöver det granskas av en områdesexpert, då oövervakad maskininläring inte själv kan avgöra om ett resultat har något egenvärde eller uppvisar kvalitet (Graham et al., 2019). Syftet med oövervakad maskininläring är att minska risken för bias (Graham et al., 2019).⁶ En alternativ användning av oövervakad maskininläring som endast har testats i en av de inkluderade studierna, är att klassificera graden av AST utifrån symptom (Minissi et al., 2019). Zhao et al. (2020) ser kombinationen av oövervakad och övervakad maskininläring som användbar vid kategorisering av rådata. En maskinell uppmärkning av rådata kan påskynda utvinningen av material till fördjupningsstudier. De benämner tekniken som Semi-Supervised Learning (SSL). Tillämpningen av SSL sker i två delfaser, en oövervakad för träning som sedan följs av övervakad finjustering (Zhao et al., 2020).

Vad är djupinläring?

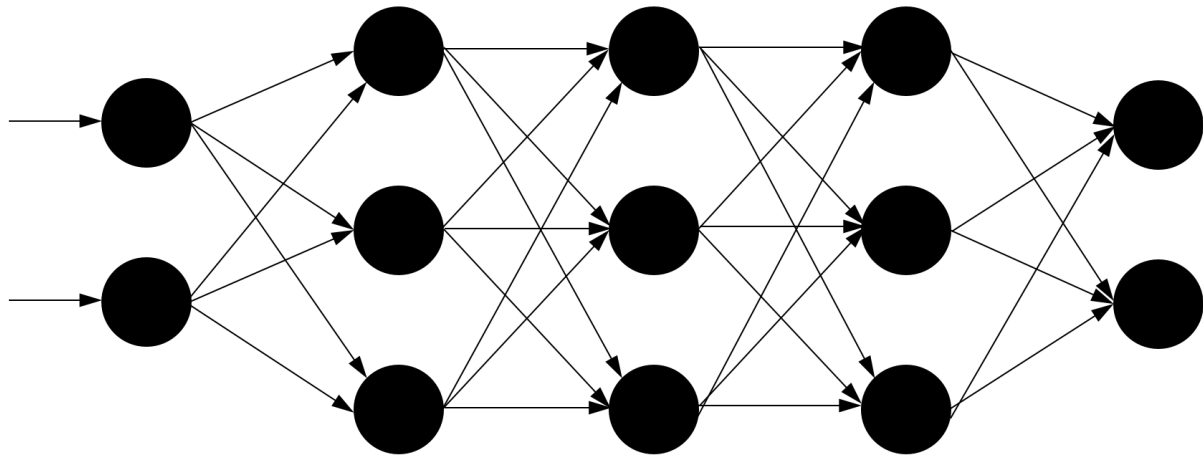
Quaak et al.:s (2021) definition av djupinläring inkluderar ett krav på två eller flera lager av dolda ekvationer mellan det inmatade värdet och resultatet, samt att de dolda lagren lär sig själv särskiljande kännetecken genom en hierarkisk inlärningsprocess. I praktiken innebär det att mellan inmatad information och svaret, alltså "källmaterialet" och "resultatet", finns dolda ekvationer som ökar komplexiteten och avser att förfina och förbättra slutresultatet i modellen. De dolda lagren utgörs av noder med koefficienter som viktat betydelsen av information, så att informationsvärdet förstärks eller dämpas. Termen dold beskriver att de kriterier som används för att bedöma och vikta information i en nod sker utan insyn eller möjlighet till påverkan utifrån. En del av bedömningen är vilken nod i nästa led som informationen förmedlas vidare till i det efterkommande lagret av noder. Varje lager använder sitt eget mått för att värdera och bedöma information (Graham et al., 2019). I det sista lagret med två noder framkommer ett resultat. Ett exempel på resultat kan vara om det

⁶ K-means algoritmen är ett exempel på oövervakad maskininläring (Kang et al., 2020a).

föreligger diagnos eller inte föreligger diagnos, där respektive alternativ är en egen nod. Figur 3 avser att illustrera sambandet mellan lager och noder vid djupinlärning, från inmatning av

Figur 3.

En illustration av olika lager vid djupinlärning.



En illustration av djupinlärning som visar ett inmatningslager (till vänster) med två noder. I mitten finns tre mellanlager med dold informationshantering, vardera med tre noder. Till höger är ett utmatningslager med två noder. Illustrationen är inspirerad av Quaak et al. (2021), men en bildsökning på "deep neural network" ger ett flertal snarlika bildträffar.

information till uppnått resultatet. Längst till vänster finns ett inmatningslager med två noder. I djupinlärningstillämpning är de lager som befinner sig mellan in- och utmatningslagret omnämna som dolda lager. Antalet dolda lager och hur många noder varje sådant lager ska bestå av förbestäms av den som sätter upp djupinlärningsmodellen, vilket kommer att påverka resultatet. För att eftersträva ett optimalt antal dolda lager och noder behöver modellen kalibreras (Misman et al., 2019). Nackdelen med de dolda lagren är att tolkningen av resultatet försvåras (Graham et al., 2019). Ett hinder i utvecklingen är att djupinlärning kräver betydligt större datamängder än maskininlärning (Zhao et al., 2020).

Förutom renodlade modeller av maskininlärning och djupinlärning förekommer kombinationer. När flera modeller av maskininlärning sammanförs blir resultatet ett "hybrid intelligent system" (Tai et al., 2019). Styrkan är en förhöjd träffsäkerhet, men den ökade komplexiteten kan även ha en negativ effekt. Det blir svårare att förstå vad som föranledde resultatet (Tai et al., 2019).

Svårigheten att få tillgång till större högkvalitativa dataset är en begränsning för utvecklingen inom AI och mental hälsa (Tai et al., 2019). Om tillgången till träningsdata är liten, väcks frågan om datamängden är tillräcklig för att hitta meningsfulla och komplexa

samband (Quaak et al., 2021). Modeller som bygger på djupinlärning med små dataset riskerar att uppvisa överanpassning till sitt träningsmaterial och ger därmed sämre generaliserbarhet på nya data (Miller et al., 2020; Zhang et al., 2021a). Zhang et al. (2021a) ser hoppfullt på modeller som kombinerar maskininlärning med djupinlärning, som de menar ofta visar bättre resultat än enskilda modeller. Samtidigt varnar Zhang et al. (2021a) för att känsligheten inom artificiell intelligens, då små skillnader kan leda till väldigt olika resultat. Det kan göra tolkning av modeller och resultat komplicerat. Tai et al. (2019) varnar för en alltför stark tilltro till att stora dataset löser kvalitetsproblem vid maskininlärning, då stora dataset påverkas av setets och algoritmens kvalitet. Datasetet och algoritmen har en likvärdig vikt för forskningsresultatet (Tai et al., 2019). Vid en granskning av studier för att diagnostisera autism med magnetic resonance imaging (MRI) fann Xu et al. (2021) en negativ korrelation mellan studiens storlek och redovisad träffsäkerhet.

Artificiell intelligens (AI) och autismspektrumstörningar (AST)

Biomarkörer har föreslagits som objektiva element för processer som sker omedvetet inne i hjärnan (Minissi et al., 2021). Bedömningar av biomarkörer kan indelas i två kategorier (Zhang et al., 2021b). Första kategorin använder en utvinningsmetod för bedömningar, exempelvis ögonskanning, elektroencefalografi (EEG) eller magnetresonanstomografi (MRI). De mest särskiljande drag som kan identifieras används för att särskilja de med och utan AST. Hur väl metoden lyckas bedöms med statistisk analys eller maskininlärning. Den andra kategorin kombinerar flera utvinningsmetoder för bedömningar. Kombinationen av information som används kan komma från biomarkörer eller vara något som bygger på ett val, som vart blicken styrs. Information som används vid utvinningen av flera metoder kan vara begränsad och kan missa att ta nytta av kombinationens fulla potential (Zhang et al., 2021b).

Idag finns inga objektiva biomarkörer för AST, i stället behöver beteendesymtom identifieras för att ställa en AST diagnos (Kojovic et al., 2021). En önskan är att skalbara metoder för utredning ska inkludera naturliga element, som hur barn med AST interagerar socialt i vardagssituationer (Kojovic et al., 2021). En skalbar metod skulle medföra en mer kvantitativ psykiatri, så att fler ges tillgänglighet för utredning utan långa väntetider. Men skalbara metoder behöver klara höga krav på precision (Washington et al., 2019).

Digitala diagnosverktyg. De första stegen för att utveckla digitala diagnosverktyg för AST kom när stabiliserade kliniska frågeformulär etablerades i praktisk verksamhet.

Stabiliseringen kom något år efter att Autism Diagnostic Observation Schedule (ADOS) och Autism Diagnostic Interview-Revised lanserats 2001 respektive 2003 (Washington et al., 2019).⁷ Ontologiskt (beskriver hur saker är) finns det tre kategorier av egenskaper som kännetecknar autismspektrum, personlighetsdrag, social kompetens och medicinsk historia (Washington et al., 2019). Digital teknik som diagnostiserar autism med artificiell intelligens vinner på att vara opåverkad av att diagnosen förändras med personlig utveckling och behandling (Washington et al., 2019). Vuxna med högfungerande autism kan ha genomgått färdighetsträning eller själva ha utvecklat strategier för att överbrygga svårigheter i social interaktion (Lio et al., 2019). Det ger belägg för att inkludera motorisk funktion vid diagnostisering av AST (Vabalas et al., 2020). Motorisk funktion kan användas som ett tidigt kännetecken som består i vuxen ålder, med digital registrering och analys med maskininlärning kan funktionen mätas direkt och vara objektivt (Vabalas et al., 2020).

Särskiljande drag och metodologiska utmaningar. Forskningen på särskiljande drag för de med AST och de utan AST kan delas in i två grenar. Den ena mäter beteenden för att finna särskiljande kännetecken, den andra utgår från bilder av hjärnan (He et al., 2021). Ett flertal studier har påvisat skillnader i blick mellan de med och utan AST, men resultaten skiljer sig betydligt och kan vara motsägelsefulla (Reimann et al., 2020). När forskningsstudier kommer med motsägelser om hur blickens rörelsemönster skiljer sig mellan individer med AST, kan blick som ett stabilt karaktärsdrag för urskiljning av personer med och utan AST ifrågasättas (Reimann et al., 2020).

Metodologiska utmaningar med att särskilja personer med AST och de utan AST med maskininlärning kan indelas i tre problem (He et al., 2021). Första problemet är att särskiljande kännetecken kodas manuellt genom efterarbete av insamlat material, vilket skapar en subjektiv bedömningsgrund (He et al., 2021). Andra problemet är användningen av maskininlärning med en binär uppdelning av personer med AST och utan AST. Här förbises att det kan finnas variation i graden av AST. Personer med AST och en låg kognitiv förmåga (under normal nivå) har ofta beteenden som är samstämmiga med sen utveckling, vilket kan ge diagnostisering på felaktiga grunder. Stereotypa beteenden som aggression och motsträvighet som förknippas med AST, kan även förekomma hos normalutvecklade individer (He et al., 2021). Tredje argumentet är att forskning som studerar hjärnabbildningar

⁷ Andra vanligt förekommande skattningsskalor för autism är Childhood Autism Rating Scale (CARS), Gilliam Autism Rating Scale (GARS) och Social Communication Questionnaire (SCQ). I tillägg används ofta Vineland Adaptive Behavior Scales (VABS), Oral and Written Language Scales (OWLS), Test of Pragmatic Language (TOPL) och Wechsler Intelligence Scale for Children (WISC) för kompletterande information (Washington et al., 2019). Men även the Social Responsiveness Scale (SRS; Minissi et al., 2021).

har inriktats på barn över 7 år samt vuxna och därmed förbises vikten av diagnos för att kunna sätta in tidiga insatser (He et al., 2021). Genom att följa och registrera ögonens rörelser med ögonskanning utifrån rörelsemönster, undersöks vilka områden blicken väljer att fokusera på (areas of interest, AOI) och vart blicken fixeras. En registrering av ”område av intresse” används för att undersöka personens spontana prioritet och mönster av uppmärksamhet (Mazumdar et al., 2021). Insamlade data om blickens visuella iakttagelser kan indikera vilken information som hjärnan processar (Kang et al., 2020a). Hur länge blicken är fokuserad på en del av en bild och var blicken stannar mätt i tidsenheter, kan användas för att visa skillnader mellan de med och utan AST. Ögonskanning kan mäta tidsskillnader med en precision på millisekunder (Zhang et al., 2021b). Skillnader i millisekunder är användbart när förmågan att ta till sig verbala instruktioner eller när socialt komplexa handlingar som kräver god kognitiv samordning mäts, det är situationer som ofta är mer utmanande för personer med AST (Zhang et al., 2021b). Personer med AST kan ha långa fixeringar vid introduktionen av en ny bild, medan de utan AST tar sig mer tid att först utforska hela bilden innan de fixerar på en del av den (Mazumdar et al., 2021). Personer med AST har en stark bias att fokusera på centrala delar av en bild, oberoende av var i bilden objekt av intresse är placerade (Mazumdar et al., 2021). Analyser av vad en bild eller video visar, används för att identifiera vad som fångar blickens intresse och kan automatiseras med djupinlärning (Li et al., 2020; Lio et al., 2019).⁸

Syfte

Studiens syfte är att med en systematisk litteraturstudie ge ett svar på frågan: Vad som krävs innan ögonskanningstekniker i kombination med artificiell intelligens kan bli ett diagnosverktyg för AST inom psykiatrisk praktik? En fördjupning i frågan tas upp i avsnittet diskussion.

Frågeställningar

- Hur används ögonskanning och datorsyn med maskininlärning för att diagnostisera AST i forskningsstudier?
- Hur används ögonskanning och datorsyn med djupinlärning för att diagnostisera AST i forskningsstudier?

⁸ Här avses deep convolutional neural networks (CNN) för bilder (Lio et al., 2019), respektive Tracking-Learning- Detection (TLD) för videomaterial (Li et al., 2020).

Metod

För att få en inblick i ett befintligt forskningsfält anses systematisk litteraturstudie vara lämplig. Ambitionen varit att följa riktlinjerna i The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 (Page et al., 2021). PRISMA 2020 består av en checklista med 27 punkter som tydliggör vilka rubriker en rapport ska inkludera, samt vilken information som ska finnas under varje rubrik.

Urval

Sökning av artiklar avgränsades till fyra medicinska/samhällsvetenskapliga databaser, APA PsycINFO, Medline, Scopus och ScienceDirect. En inkludering av tekniskt inriktade databaser skulle kunna ha givit ett bredare urval av forskning om artificiell intelligens i kombination med ögonskanning, än det som har inkluderats här.

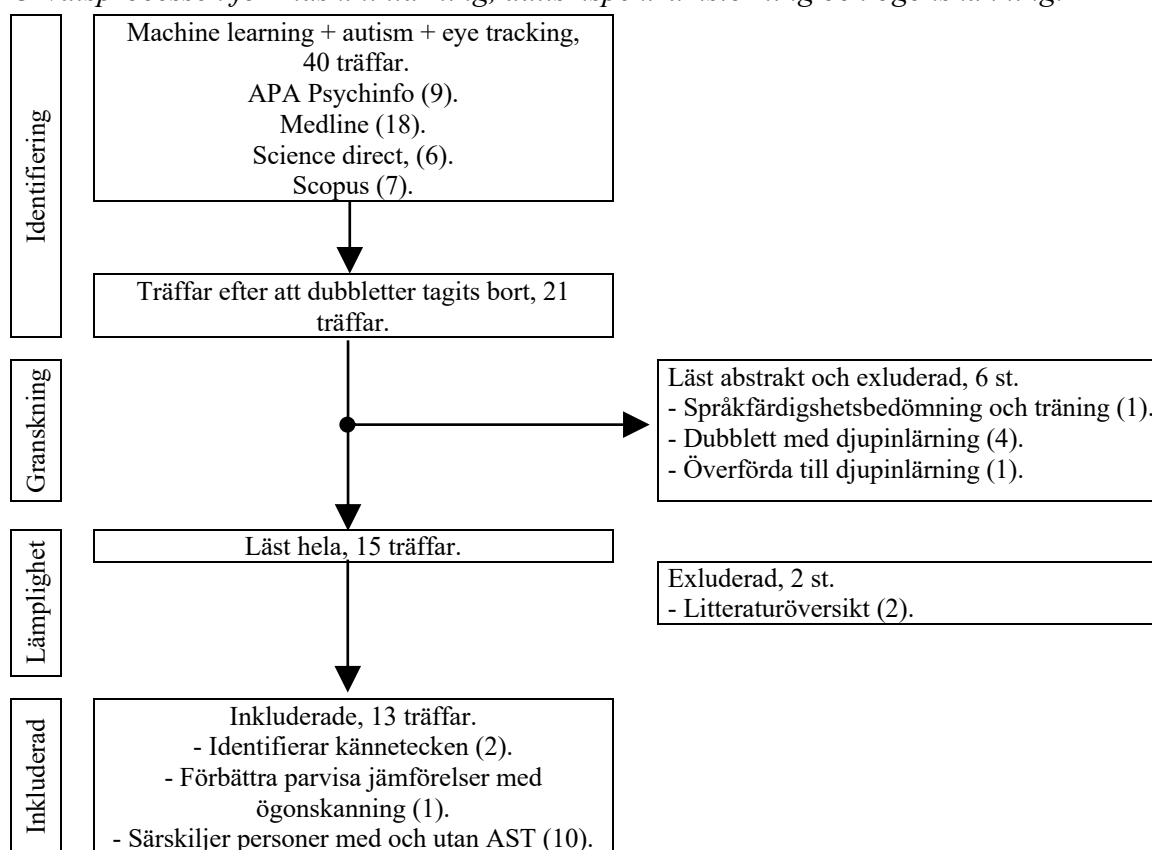
För att begränsa omfånget har forskning som har en annan avsikt än psykiatrisk diagnostik valts bort. Artiklar inkluderade i studien ska bidra till området psykologisk eller psykiatrisk diagnostik, och uttryckligen använda artificiell intelligens i formen av maskininlärning eller djupinlärning för diagnostik av AST med ögonskanning.

Kriterier för sökningen är att artiklarna ska vara tillgängliga via Lunds universitet, publicerade i vetenskapligt granskade journaler (peer-reviewed), publicerade på engelska och vara inkluderad i någon av nämnda databaser inom perioden 2019-01-01 - 2021-11-30. En sista uppföljande kontroll av sökresultatet gjordes 2022-01-07. Valet att endast gå tillbaka två år i tid motiveras med att det är den senaste forskningen som antas vara närmast kliniska tillämpningar. Ingen oberoende granskare har kontrollerat sökresultatet, utan artikelgranskning och urval har endast utförts av författaren till uppsatsen.

Första söksträngen var ”machine learning + autism + eye tracking”, vilket gav 21 träffar, figur 4.

Figur 4.

Urvalsprocessen för maskininlärning, autismspektrumstörning och ögonskanning.



I rutan med inkluderade artiklar framkommer en total på 13 träffar, vilka har delats in i tre underkategorier. De två artiklar som ”identifierar kännetecken” undersöker vad som särskiljer individer med och utan AST. Orden ”identifierar kännetecken” betonar att fokus är ett explorativt undersökande. Avsikten är att se vilka metoder som framtida studier kan använda, de testar inte hur väl framkomna kännetecken fungerar i praktiken.

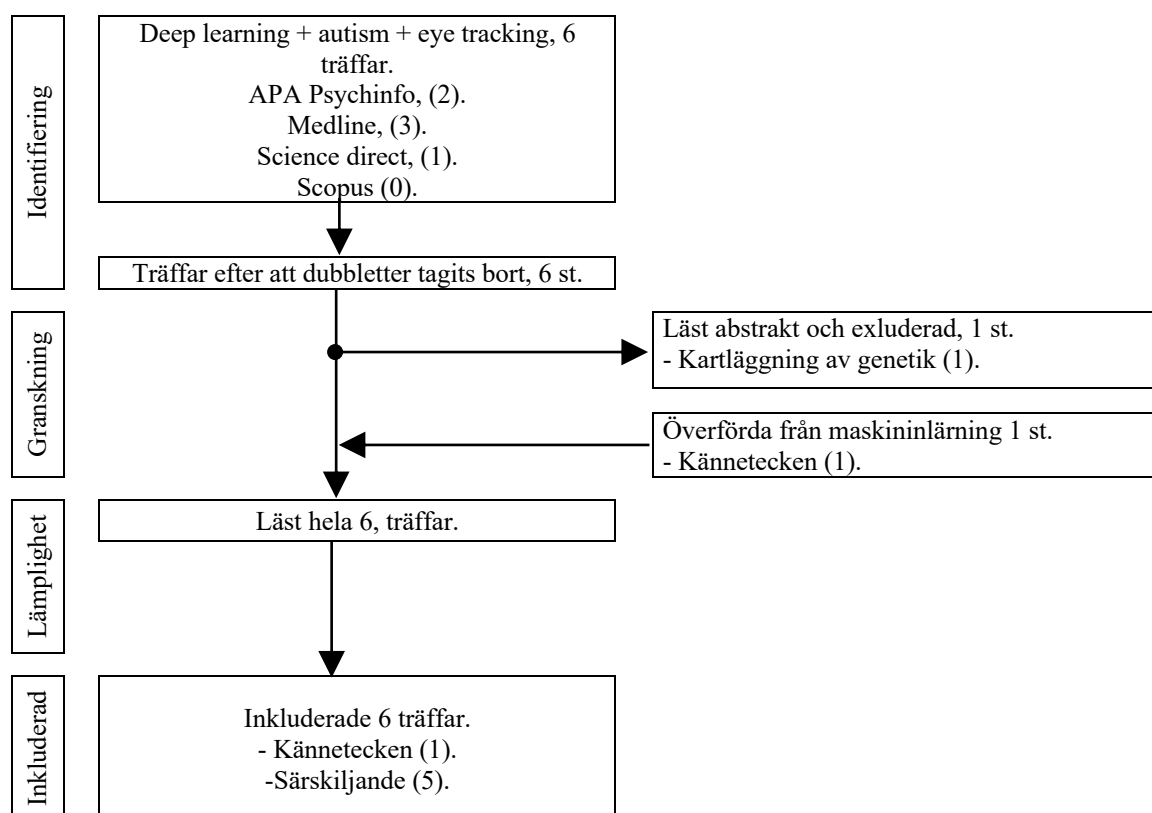
En artikel har som mål att ”förbättra parvisa jämförelser med ögonskanning”, där målet är att utvinningen av data från studier ska förbättras.

10 artiklar ”särskiljer personer med och utan AST” och inom den här kategorin kan två syften identifieras. Det första syftet är övergripande, då de olika studierna använder olika kombinationer av metoder för att särskilja om en individ har AST eller inte. Målet är att finna en tillämpning som uppnår högsta möjliga träffsäkerhet, samtidigt som antalet felaktiga diagnostiseringar minimeras. Det andra syftet är att inom varje enskild studie undersöka vilken algoritm som ger högsta träffsäkerhet och minst antal fel. Varje studie redovisar ett resultat för två eller flera algoritmer. Här kommer fokus vara på den högsta träffsäkerhet som varje studie uppvisar.

Den andra söksträngen var ”deep learning + autism + eye tracking” 6 träffar, figur 5.

Figur 5.

Urvalsprocessen, deep learning + autism + eye tracking.



Totalt två artiklar har haft ett annat syfte och uteslutits efter en genomläsning av sammanfattningen. En avsåg förbättrad färdighetsträning av språk för barn med autism och framkom i urvalet för maskininläring. Den andra artikeln hade syftet att kartlägga gener som påverkar social uppmärksamhet och framkom i urvalet för djupinläring.

Fyra artiklar var dubbel träff i de båda sökningarna och en artikel flyttades från maskininläring till djupinläring efter valet av algoritmer framkommit. Litteraturoversikter har lästs i sin helhet, men har exkluderats från resultatdelen. De två litteraturstudier som framkommit var Minissi et al. (2021) och Washington et al. (2019).

Nedladdade artiklar har sparats med första författarens efternamn och artikelns publicerade årtal. Varje fil sorterats i hierarkiska system av mappar. Det har varit en mapp för varje huvudkategorierna, maskininläring och djupinläring. Inom en kategori flyttades artiklar från "olästa" till "lästa och inkluderade", "läst sammanfattning och exkluderad", "läst hela och exkluderad", "läst review artikel" eller "dubbelträff". De inkluderade artiklarna indelades i "särskiljande studier" eller "undersöker kännetecken". Om en artikel blev avvisad,

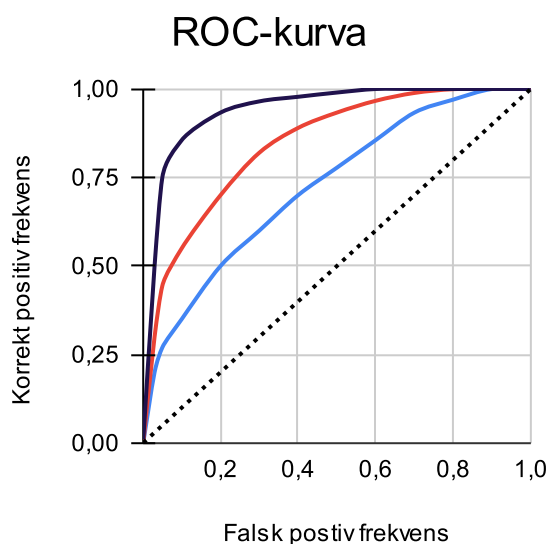
placerades den i en undermapp med ett namn som tydliggjorde skälet. Hanteringen möjliggjorde att processen kunde granskas i efterhand. Risken för bias i urvalet bedöms som låg, då de två artiklar som exkluderats utöver litteraturoversikter har varit tydligt skilda från uppsatsens syfte och mål. I tabeller har avsaknad av uppgift markerats med ett streck (-).

Bedömningskriterier för AI

Graham et al. (2019) presenterar de mest frekvent förekommande bedömningskriterierna när AI används för att beräkna sannolikheten för ett visst utfall, the Area Under Curve (AUC),⁹ träffsäkerhet (accuracy), sensitivitet,¹⁰ specificitet¹¹ och precision.¹² Mindre frekventa förekommande mått är F1-värde,¹³ precision.¹⁴ Receiver operating characteristics (ROC) är en grafisk redovisning av testmetoders prestanda (figur 6).

Figur 6.

Receiver operating characteristics (ROC).



Exempel på receiver operating characteristics (ROC). Den area som under linjen benämns som Area Under Curve (AUC). För AOC är lägsta möjliga värde från .00 och högsta värde är 1.00. Den sträckande linjen motsvarar slumpmässig klassificering, dvs. 50% korrekt bemöda värden och 50% där bedömningen är felaktig. En kurva som visar en hög grad av korrekta och få felaktiga bedömningar får en linje som går nära det övre vänstra hörnet.

Om resultatet från ett test ska anses positivt och personen bör tillskrivas diagnos, beror på var tröskelvärdet för att tillskriva diagnos sätts. Illustrerat som en graf visar Y-axeln

⁹ Area Under Curve (AUC) är ett prestandamått vid binär klassificering.

¹⁰ Sensitivitet är synonymt med sant positivt värde/true positive rate (TPR) och recall (Graham et al., 2019).

¹¹ Specificitet är synonymt med sant negativt värde/true negative rate, (NPR; Graham et al., 2019).

¹² Precision är synonymt med positivt prediktivt värde/positive predictive value (PPV; Graham et al., 2019).

¹³ F1-värde = $2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$.

¹⁴ Precision = totalt antal korrekta identifierade värden med diagnos / hela gruppen med diagnos.

korrekt positiv frekvens (True Positive Rate, TPR), alltså den del av populationen som algoritmen tillskrivit en diagnos och verkligen har diagnosen dividerat med totalt antal positiva.¹⁵ X-axeln visar falsk positiv frekvens (False Positive Rate, FPR), det är antal individer i populationen som algoritmen har tillskrivit en diagnos, men inte anses uppfylla kraven för diagnos utan borde ha kategoriserats som ”frisk”, dividerat med totalt antal friska.¹⁶ Det gränsvärde som väljs är avvägning om sannolikheten för över respektive underdiagnostisering.

Underdiagnostisering uttrycks som falsk negativ frekvens (False Negative Rate, FNR) och är när diagnos finns hos individen, men individen missas av instrumentet.¹⁷ Själva ROC-kurvan framkommer i en plottad graf för varje tröskelvärde och kurvan visar testets sensitivitet (Statens beredning för medicinsk och social utvärdering, 2014).

Arean under ROC-kurvan omnämns som AUC. När resultat för $AUC = 1$ kan algoritmen urskilja grupper korrekt. När resultatet är lika med .5 har algoritmen ingen förmåga att urskilja grupper, då korrekta och felaktiga resultat överlappar varandra. När $AUC = 0$ innebär att alla föreslagna kategoriseringar är felaktiga (Graham et al., 2019). Resultatmässigt anses $AUC > .7$ acceptabelt och $AUC > .8$ anses som excellent (He et al., 2021).

Träffsäkerhet beräknas genom att addera korrekta positiva och negativa träffar till en ”summa av korrekta värden”. Summan av korrekta värden divideras med summan av ”korrekt positiva, korrekt negativa, falskt positiva och falskt negativa värden”.¹⁸ Rapportering av träffsäkerhet kan anges i procent eller som decimaltal. Måttet träffsäkerhet anses olämpligt att använda om det är stor skillnad i fördelningen mellan grupp-kategorier, exempel få personer med AST och stor kontrollgrupp (Graham et al., 2019).

Sensitivitet kan ibland omnämnas som korrekt träffsäkerhetsvärde (True positive rate, TPR) eller ”recall”. Värdet för sensitivitet avser visa andelen korrekt positiva träffar (True Negative Rate, TNR) som identifieras av alla personer som var inkluderade i en kategori.¹⁹ Specificitet är motsvarande värde för antalet korrekt negativa träffar.²⁰ Ofta interagerar de två värdena, så att en ökad sensitivitet ger en försämrad specificitet och omvänt (Graham et al., 2019).

¹⁵ Korrekt positiv frekvens (TPN) = korrekt positiva / (korrekt positiva + falskt negativa).

¹⁶ Falsk positiv frekvens (FPR) = falskt negativa / (falskt negativa + falskt positiva; Mazumdar et al., 2021).

¹⁷ Falsk negativ frekvens (FNR) = falskt negativa / verkligt antal med diagnos (Mazumdar et al., 2021).

¹⁸ Träffsäkerhet = (korrekt positiva + korrekta negativa) / (korrekta positiva + korrekta negativa + falskt positiva + falskt negativa).

¹⁹ Sensitivitet = korrekt positiva / (korrekt positiva + falskt negativa).

²⁰ Specificitet = korrekt negativa / (korrekt negativa + falskt positiva).

Resultat

Redovisningen kommer att ske i turordning och börjar med studier som endast tittar på kännetecken. Efter följer särskiljande studier för personer med och utan AST. Resultatet från maskininlärning kommer att följas av studier för djupinlärning. Vid slutet av varje avsnitt kommer det att finnas en tabell över inkluderade deltagare. För de särskiljande studierna redovisas en sammanställning av studiernas resultat. Användningen av underrubriker i resultatdelen används för att indikera var sammanfattningen av en ny studie tar vid. Namnet i underrubriken avser att ge läsaren en vägledning om något som kännetecknar den studiens innehåll.

Studier som inriktar sig på kännetecken och använder maskininlärning

Observation genom datorsyn. Sapiro et al. (2019) har valt att redovisa tre delstudier med småbarn som målgrupp. Gemensamt för alla tre delstudierna är att de har använt datorsyn. Urvalet är en mindre del i en större studie med 1756 deltagare och 4441 videos (Sapiro et al., 2019). De tre delstudierna är genomförda med syftet att finna tillförlitliga kännetecken för att särskilja barn med och utan AST. Endast i en av delstudierna framkommer ålder, men forskargruppens fokus är på småbarn. Det långsiktiga målet är att utveckla skalbara verktyg för grovgallring (screening) med datorsyn. Fördelen med datorsyn är att tekniken är enkel att tillämpa, och det räcker med hemelektronik som smarta telefoner eller surfplattor för att utföra enkla tester. På skärmen visas stimuli som används för att utvärdera iakttagarens intresse, parallellt kan kameran registrera ögonrörelser. Insamlade data processas med maskininlärning (Sapiro et al., 2019).

Sapiro et al. (2019) nästa delstudie inkluderar 108 barn ($n = 22$ med AST och $n = 86$ utan AST). Ålder framkommer inte. Det som studeras är vart uppmärksamheten riktas när barnet sitter framför en skärm som visar två skilda stimuli. På ena halvan av skärmen visas en kvinna som sjunger (socialt stimulus) och på den andra halvan en teddybjörn (osocialt stimulus). Var 60 sekund skiftar den sjungande kvinnan och teddybjörnen plats. Barnen med AST visade generellt mindre uppmärksamhet för videon. Det förekom att huvuddelen av tiden fokuserade de på icke sociala stimuli, vilket var ovanligt i kontrollgruppen. Barn med AST fixerade i högre utsträckning sitt fokus på en av de stimuli som visades på skärmen (social eller icke social). Sapiro et al. (2019) ser valet av socialt/osocialt stimuli, graden av uppmärksamhet på videon och fixering på självvalda stimuli som tre potentiella biomarkörer för att särskilja de med och utan AST.

Små barn med AST intar oftare asymmetriska armpositioner och stereotypa motoriska rörelser (Sapiro et al., 2019). Därför utformades en andra delstudie för digitalt iakttagande av kroppsrörelser med datorsyn när barnet tittar på en video. Data från 104 barn samlas in (n = 22 med och n = 82 utan AST, ålder 16-31 månader). För att registrera barnets huvudrörelser hade en kamera placerats ovanpå huvudet. När barnet var upptaget att titta på en film uttalades deras namn av en person bakom barnet. Om de reagerade och hur snabbt de eventuellt vred på huvudet fångades med kameran. För barnen med AST var det 8% som vred på huvudet för att besvara anropet. I kontrollgruppen var det 63% av barnen som reagerade. Barnen med AST hade en längre genomsnittlig responstid (2.02 vs 1.06 s, $p = .04$ [Sapiro et al., 2019]).

Sapiro et al. (2019) tredje delstudie utgår från att barn med neurologiska sjukdomar kan ha försämrade förmåga att uttrycka sina känslor via ansiktsmimik. Målet med undersökningen var att skanna av barnets ansikte med datorsyn och bedöma uttryckta känslor utan mänsklig inblandning. Om datorsyn registrerar känslouttryck är det påtagligt enklare att genomföra tester i större skala, än när människor övervakar och registrerar data manuellt. För den här delstudien framkommer inte hur många deltagare som ingår eller vad deras ålder är. Resultaten visar träffsäkerhet 0,89, sensitivitet .90 och F1-värde .89 (Sapiro et al., 2019).

Igenkänning av känslor. Król och Król (2019a) ville undersöka om personer med AST har sämre förmåga att avläsa känslor. Studien har 44 deltagare (n = 21 med AST, $m = 16.27$ år, $SD = 4.84$ år och n = 23, $m = 16.31$ år, $SD = 2.69$ år) och spänner över åldersspannet, 12-30 år.²¹ Var i åldersspannet som en individ befinner sig kan ha en stor betydelse för personlig mognad. För att undvika att resultatet påverkas av den typen av olikheter, är deltagarna matchade utifrån kön och ålder (Król & Król, 2019a).

Ögonskanning med maskininlärning används i studien för att se var blicken fixeras när deltagarna bedömer 36 ansiktsbilder. Instruktionerna skilde sig för varje serie om 12 bilder, att endast titta på ansiktet, specifikt bedöma personens känsla, samt att bedöma vilken av munnen eller ögonbrynen som är tjockast. Inför vardera av de tre omgångarna med 12 bilder gavs deltagarna olika instruktioner om vad de skulle iaktta. Oavsett vad instruktion var, så var forskarnas mål att se hur väl deltagarna klarade att bedöma vilken känsla personen på bilden visade. Personer med AST hade signifikant fler felbedömningar. Generellt fixerade de blicken mellan ögon och mun oberoende av vad de uppmanades att titta efter. Personer utan AST visade på förändringar i sin strategi, när uppmaningen var att bedöma känslan som

²¹ SD = standardavvikelse för ålder.

personen uppvisar i bilden. Król och Król (2019a) menar att en del personer med autisms svårighet att avläsa känslor delvis beror på en sämre förmåga att läsa av ansikten.

En sammanställning över antalet deltagare för de två studierna finns i tabell 1.

Tabell 1.

Översikt av studier som använder maskininlärning med fokus på kännetecken.

| Författare | Inriktning på studien | Inkluderade | Antal deltagare | Med ASD | Medel ålder | Antal flickor | Utan ASD | Medel ålder | Antal flickor |
|-----------------------|---|------------------------|-----------------|---------|--------------------|---------------|----------|--------------------|---------------|
| Sapiro et al. (2019)* | Datorsyn, intresse för social respektive osocial stimulus | Småbarn | 108 | 22 | - | - | 86 | - | - |
| | Datorsyn, huvudrörelser | Småbarn, 16-31 månader | 104 | 22 | - | - | 82 | - | - |
| | Datorsyn, affektanalys | Småbarn | - | - | - | - | - | - | - |
| Król & Król (2019a) | Ögonskanning, identifiera känslor | 12-30 år | 44 | 21 | 16.27 SD = 4.84 | 2 | 23 | 16.31 SD = 2.69 | 5 |

SD = Standardavvikelse.

*Hela studien inkluderar 1756 deltagare och 4441 videos. Artikeln redovisar endast de delar av studien som använt datorsyn. För delstudien affektanalys anges inget deltagarantal.

Studie som avser kännetecken och använder djupinlärning

Sekvenser ur Hollywoodfilmer. Inom djupinlärning var det endast Reimann et al.:s studie som fokuserade på kännetecken (2020). Alla 69 deltagare var män och matchades i ålder och IQ. En grupp var högfungerande AST ($n = 33$, $m = 20.25$ år, $SD = 4.02$ år) vilka jämfördes med en grupp utan AST ($n = 36$, $m = 20.83$ år, $SD = 3.82$ år [Reimann et al., 2020]).

Noterbart är att totalt 50 deltagare med ASD antogs till studien. 17 av dem är inte med i analysen. Orsaker till bortfallet var att deltagarna inte fullföljde hela testet ($n = 5$), kvaliteten på datainsamlingen blev för låg ($n = 4$), kvalificerade inte för AST diagnos ($n = 3$), svårigheter med tidsbokning ($n = 2$), fullt IQ test visade en nivå under 70 poäng ($n = 1$), konflikter på grund av medicinering ($n = 1$) och missad uppföljning ($n = 1$ [Reimann et al., 2020]).

Studien avsåg att kontrollera om blickens placering är ett stabilt särskiljande kännetecken. Varje deltagare satt med huvudet fixerat och med en ögonskanner på höger öga. De visades 14 sekunders klipp ur 24 Hollywoodfilmer och avsikten var att se om det fanns skillnader i preferenser för vart blicken riktades. För varje visad filmsekvens hade en artificial

neural network (ANN) algoritmer identifierat områden av intresse.²² I analysen undersöktes om det fanns tidsskillnader i blickens iakttagande av ögon, näsa, mun respektive om blicken var fokuserad på ”ansiktet” eller ”något annat än ansiktet”. Resultatet visade att individuella preferenser om vart blicken var placerad stabiliserades när fler filmsekvenser ingick i analysen. Gruppen med AST tittade generellt mindre på det centrala ansiktets område vid näsan och mer på ”något annat än ansiktet” än personerna utan AST. Grupperna hade inga signifikanta skillnader för tiden då blicken var fokuserad på ögon och mun (Reimann et al., 2020). Reimann et al. (2020) tror att en orsak till att resultatet för ögon och mun skiljer sig från tidigare forskning, är att de i regel har utgått från stillbilder i stället för rörliga bilder. Variationen i vart blicken var fäst var större inom gruppen med AST än den utan AST. Det tyder på att mängden data behöver vara större för att korrekt bedöma personer med AST och att för lite data kan ge felaktiga svar (Reimann et al., 2020). En översikt för deltagarna finns i tabell 2.

Tabell 2.

Översikt av studie som använder djupinlärning med fokus på kännetecken.

| Författare | Inriktning på studien | Inkluderade | Antal deltagare | Med ASD | Medel ålder | Antal flickor | Utan ASD | Medel ålder | Antal flickor |
|-----------------------|---|-------------|-----------------|---------|--------------------------|---------------|----------|--------------------------|---------------|
| Reimann et al. (2020) | Ögonskanning, preferenser för iakttagande av film | 12-30 år | 69 | 33 | 20.25 år SD = 4.02 år | 0 | 36 | 20.82 år SD = 3.82 år | 0 |

SD = standardavvikelse för ålder.

Studier som särskiljer AST / inte AST med maskininlärning

I de inkluderade studier som särskiljer personer med och utan AST med maskininlärning är det en stor spridning på deltagarnas ålder. En studie fokuserar på småbarn (Chang et al., 2021). Fyra studier har huvudsakligen inkluderat förskolebarn (Alcañiz et al., 2021; He et al., 2021; Kang et al., 2020a; Kang et al., 2020b; Wan et al., 2019). Tre studier har en variation av skolbarn med en spridning från 5-17 år (Mazumdar et al., 2021; Nag et al., 2020; Zhao et al., 2021) och en studie utgår från vuxna deltagare (Vabalas et al., 2020). När studierna redovisas har ordningen försökt spegla valet av metod, i stället för deltagarnas ålder. Avsnittet har även en studie som har fokus på metodförbättring.

Objektets placering vid utforskande av stillbilder. Mazumdar et al. (2021) har undersökt hur bildinnehåll och sättet att iaktta bilden kan kombineras för att urskilja barn med

²² Artificial neural network (ANN) är en algoritm för djupinlärning.

och utan AST. Syftet var att undersöka om objektets placering i bilden har betydelse för betraktarens fixeringspunkter och centrumbias. Centrumbias är en variabel som inte har använts lika tydligt i övriga inkluderade studier och som undersöks med ögonskanning. Det finns flera sätt att skapa variabler för centrumbias. Ett sätt är att beräkna ”områden av intresses” avstånd till mitten av bilden. Andra variabler är medelavståndet för alla områden av intresse, det totala antalet fixeringspunkter i centralområdet och totaltiden då blicken är nära bildens mitt (Mazumdar et al., 2021).

Mazumdar et al.:s (2021) studie utgår från ett konferensmaterial där 28 barn ($n = 14$ med AST och $n = 14$ utan AST, 5-12 år) har tittat på 400 bilder.²³ Undersökandet av särskiljande algoritmer är omfattande, 23 olika varianter. Det högsta resultatet är i linje med andra studiers publicerade resultat på samma material, träffsäkerhet .59, sensitivitet .68, specificitet .50 och F1 värde på .61, precision .57 och det finns utrymme för förbättringar (Mazumdar et al., 2021).²⁴ En skillnad från studier som använt samma dataset är hur Mazumdar et al. (2021) väljer att grafiskt redovisa sitt resultat. De använder ett värmekamerafilter. På bilden nyanseras ett områdes färg beroende på graden av uppmärksamhet det fångat. Med ökad grad av uppmärksamhet blir färgen varmare och intensivare. Beroende på bild, så syns skillnaderna mellan grupperna med varierande tydlighet. Antalet kännetecken som används för i det redovisade resultatet framgår inte.

Datorsyn och uppmuntrande ledtrådar. He et al. (2021) har genomfört en automatisk studie med datorsyn med 74 förskolebarn matchade i ålder och indelade i tre kategorier. En illustration från studien visades när ögonskanning skulle förklaras (figur 2, sid 4). Första kategorin är högfungerande med AST ($n = 26$, $m = 5.08$ år, $SD = 0.83$ år, 23 pojkar och 3 flickor), andra kategorin är lågfungerade med AST ($n = 24$, $m = 4.98$ år, $SD = 1.09$ år, 19 pojkar och 5 flickor), tredje kategorin är barn utan AST ($n = 24$, $m = 5.24$ år, $SD = 0.81$ år, 18 pojkar och 6 flickor [He et al., 2021]). På en skärm visas fem bilder, en bild i mitten och en på varje sida av den centrala bilden. I den centrala bilden indikeras att barnet ska rikta sin blick till en av de yttre bilderna och bibehålla blicken där. Om det lyckas får barnet uppmuntran. Om barnet misslyckas, så händer inget. Själva indikationen ges med fyra olika typer av stimuli (He et al., 2021). Ett stimulus är ”ett mänskligt ansikte” som med ögonen tittar mot en av sidobilderna och ger barnet vägledning vart det ska föra blicken. Övriga stimuli är ”ett tecknat ansikte” som visar riktning med ögonen, ”en hand” som pekar samt ”ett

²³ Datasetet är från Integrated computational materials engineering, ICME Grand challenge.

²⁴ Andra redovisade mått är precision .57, andelen falskt positiva .49 och andelen falskt negativa .32 (Mazumdar et al., 2021). Det var den enda studien som redovisade resultat för falskt positiva och falskt negativa träffar.

piltecken” som riktas mot någon av sidorna. Barnen utan diagnos klarar uppgiften bäst. Barn med högfungerande autism något sämre. De som klarade uppgiften sämst är barnen lågfungerande autism. Hur blicken har rört sig (spatialt) blir den faktor som används för att särskilja grupperna med en maskininlärnings algoritm.²⁵ Träffsäkerhet för att särskilja de tre grupperna är 81.08% och inkluderade alla 68 kännetecken som användes i studien (He et al., 2021). Det vanligaste felet vid urskiljning är att någon av de utan AST diagnos eller med lågfungerande AST hamnade i gruppen för högfungerande AST. AUC är redovisat per grupp, högfungerande autism .775, lågfungerande autism .908 och barn utan AST .927 (He et al., 2021). När de båda AST kategorierna behandlas som en grupp ökade träffsäkerheten till 93.24% och AUC .938 (He et al., 2021). Antalet kännetecken som inkluderats anges inte. Det som artikeln lyfter fram som en styrka är att hela experimentet tar 15 minuter per undersökning och är lätt att administrera (He et al., 2021).

Etnisk ansiktsigenkänning. Kang et al. (2020a) i sin studier 157 barn (3-6 år). En grupp är har AST och låg begåvning (IQ < 70, n = 77) och den andra gruppen är normalbegåvade barn utan diagnos (IQ > 90, n = 80). Studien utgår från ansiktsigenkänning och undersöker vad som är mest särskiljande för grupperna genom att granska vart på en bild barnen tittar när de ska avgöra om personen har en etnicitet som barnet själv. Bilderna kan föreställa bekanta eller obekanta ansikten med samma etnicitet, personer med en annan etnicitet eller en mix av de tre grupperna. Vart barnen fäster blicken dokumenteras med ögonskanner. Den högsta träffsäkerheten 84.17% i att särskilja barnen med AST och låg begåvning från gruppen utan AST är när alla bilder inkluderas, AUC värde .89 (Kang et al., 2020a). Algoritmen som särskilde grupperna åt tilläts att använda 120 olika kännetecken för uppgiften (Kang et al., 2020a). Barn med AST hade ökad fokus på kinder, bakgrund och kroppsdelar, vilka ger mindre information om sociala kännetecken om den som betraktas. Barnen utan AST fokuserade blicken mer på ögon, näsa och mun vilket ger rikligare information om sociala aktiviteter (Kang et al., 2020a).

Etnisk ansiktsigenkänning och elektroencefalografi (EEG). Elektroencefalografi (EEG) är en metod som mäter och registrerar hjärnaktivitet genom elektroder. En undersökning tar ca 6 minuter (Kang et al., 2020b). Metoden har hög tidsupplösning och har visat sig vara ett kraftfullt verktyg att använda vid komplexa neuropsykiatriska diagnoser (Kang et al., 2020b).²⁶ Forskning som mäter hjärnaktivitet hos barn med normalbegåvning

²⁵ De använder K-nearest neighbors algorithm, en icke-parametrisk klassificeringsmetod.

²⁶ Känsligheten i mätningarna gör att responsen i hjärnaktivitet vid ett förändrat stimuli kan mätas i millisekunder (Müller et al., 2020).

har kunnat påvisa ett tydligt utvecklingsmönster. Med ökande ålder sjunker antalet rytmer med låg frekvens i hjärnan, samtidigt ökar i motsvarande grad antalet rytmer med hög frekvens. För barn med AST har styrkan i nämnda signalerna visat karaktäristiska skillnader jämfört med normalutvecklade barn. Skillnader i signalstyrka har potential att användas som kännetecken för att särskilja barn med AST vid diagnostisering (Kang et al., 2020b). En undersökning med 97 barn (ålder 3-6 år, med AST, n = 49, utan AST, n = 48,) genomfördes. En klassificering baserad på styrkan i EEG signaler gav en träffsäkerhet på 68%. I ett separat deltest hade barnen som uppgift att titta på porträttbilder av ansikten av främmande människor och identifiera om de tillhörde samma eller annan etnisk grupp än de själva. Vart på bilden barnen fokuserade blicken och hur länge varje blick förblev inom varje delområde registrerades. Indelningen bestod av åtta kategorier, bildens bakgrund, personens bål, ansikte, ögon, höger öga, vänster öga, mun och näsa. Träffsäkerheten för att urskilja de med AST och utan AST var högre, 75.89% än vid EEG studien. Kombinationen av EEG och igenkänning av ansikten visade på den högsta träffsäkerheten 85.44%, antalet inkluderade kännetecken var 32 (Kang et al., 2020b).

Identifiering av känslor i ansiktet. Nag et al. (2020) prövar användning av smarta glasögon för att se om man kan skilja på 33 ungdomar (6-17 år) med och utan AST. Ungdomarna har i uppgift att identifiera en känsla per porträttbild. De ska avgöra om personen på bilden uppvisar glädje, ledsenhet, ilska, rädsla, äckel, överraskning eller lugn. De noterar att de med AST ägnar mindre tid till varje bild, men resultatet är inte signifikant och resultatet redovisas inte. Slutsatsen de drar är att ungdomar med AST som ska göra sociala bedömningar, är mer benägna att fokusera på distraherande stimuli än de utan AST (Nag et al., 2020).

Video med munrörelser. Wan et al. (2019) har i sin studie inkluderat 74 barn, hälften med AST (n = 37, m = 4.6 år, SD = 0.7 år, 33 pojkar och fyra flickor) och hälften utan AST (n = 37, m = 4.6 år, SD = 0.4 år, 27 pojkar och 10 flickor). Ögonskanning med maskininlärning har använts för att undersöka ”områden av intresse” när barnen visade en 10 sekunders video. Videon var utan ljud och visade ett porträtt av en kvinna som bokstaverar det engelska alfabetet. Barnen med AST tittade signifikant mindre på ögon, mun, ansikte, kropp och hår. Det räckte att se på hur barnen iakttog mun och kropp för att skilja på grupperna, vilket gav träffsäkerhet 85.1%, sensitivitet 86.5% och specificitet 83.8% (Wan et al., 2019). Antalet kännetecken redovisas inte.

Mellanmänsklig interaktion med ögonskanning. Vanligen genomförs studier med ögonskanning vid uppspelning av video. För att utesluta att resultatet inte skiljer sig vid

verklig interaktion med en annan människa, behöver en annan typ av experiment genomföras. För att säkra att barnen har goda färdigheter i kommunikation testade Zhao et al. (2021) i sin studie, 39 barn (6-13 år, $n = 20$ med AST och $n = 19$ utan AST). De har på förhand screenats som normalbegåvade. Ytterligare fyra deltagare var antagna till studien, men de föll bort av på grund av tekniska problem (bortfall med AST $n = 1$, utan AST $n = 3$). Barnen blev intervjuade och fick ställa frågor när de satt mitt emot en intervjuare. Intervjuaren var omedveten om vilken grupp barnen tillhörde. Frågorna till barnen var fördelade på fyra sessioner, generella frågor, vilken hobby, ja/nej frågor och i sista sessionen fick barnen ställa frågor till intervjuaren. Barnen hade glasögon som skannade barnets blick och de instruerades att titta på intervjuaren och undvika plötsliga rörelser. Vart blicken var fokuserad registrerades som tidsandel fördelat på fyra kategorier, intervjuarens kropp, hela ansikte, mun eller ögon. Ytterligare kännetecken som användes för att undersöka skillnader mellan barnen var totaltiden för varje delsession, samt totaltid för alla fyra sessioner. Den högsta uppmätta träffsäkerheten var 92.31%, som uppnåddes med en kombination av visuell fixering av blick och sessionslängd, specificitet 100%, sensitivt 84.21% och $AUC = .92$ och tre inkluderade kännetecken.²⁷ Den högsta träffsäkerheten för enbart blick var 84.62%, samma träffsäkerhet uppnåddes med enbart sessionslängd 84.62% (Zhao et al., 2021).

Den tydligaste skillnaden mellan grupperna var att de med AST spenderade en högre andel av totaltiden med blicken fokuserad på mun respektive hela kroppen. Men även om antalet frågor i varje session var lika fördelat för alla deltagare, tog intervjun i regel längre tid för barn med AST. Studien exkluderade barn med ADHD och sen utveckling (Zhao et al., 2021).

Datorsyn och social preferens. En omfattande studie av Chang et al. (2021) inkluderade 993 småbarn ($m = 21.1$ månader, 17.1-36.9 månader). Barnen var indelade i tre grupper, 40 barn hade fått remiss för utredning av AST, 17 barn hade kategoriserats som en riskgrupp på grund av sen utveckling eller sent utvecklat språk, men nådde inte kriterierna för AST, 936 barn ansågs följa en normal neurologisk utveckling. Studien har redovisat könsfördelning för alla inkluderade, men saknar precisering i undergrupperna.

Barnen fick i sällskap av sin vårdnadshavare under en 10 minutersperiod se ett par kortare filmsekvenser. Två av filmerna undersökte social preferens, hur mycket tid fokuseras på att se en människa som leker med en leksak, alternativt på en distraherande leksak på den andra halvan av skärmen. En av barngrupperna fick även se en film som undersökte intresset

²⁷ Resultatet uppnåddes med en support vector machine (SVM).

för växlande konversationer. På skärmen visas en konversation mellan två individer, en vid vardera bildkanten. Testet undersökte om barnets blick följde med när talet växlade mellan individerna. Det fanns två kontrollfilmer som varken innehöll människor eller prat och inte har inkluderats i AST bedömningen. Filmerna visades antingen en smart telefon eller en läsplatta. Inläsning av ögonrörelser gjordes med datorsyn i samma applikation som används för att spela upp filmerna. För utvärdering matchades barnen med AST med barn från kontrollgruppen i samma ålder. Skillnader prövades statistiskt för på gruppnivå och för att undersöka individuella skillnader användes klassificeringsträd baserat på maskininläring. Barnen med AST uppvisade mindre koordination gällande var de fäste blicken vid växlande konversationer, det framträdde även skillnader i förmågan att följa social interaktion. Vid utvärdering var träffsäkerheten högre när data från två filmer, AUC .88 eller tre filmer AUC .90 sammanvägdes, än för enskilda filmer AUC .76-.84. Studien uppger att många kännetecken har inkluderats, men säger inte hur många (Chang et al., 2021).

Förutom att vara ett enkelt sätt att screena barn till en låg kostnad, har metoden möjligheten att testa barn i tidig ålder. De beteenden som mättes utvecklades under barnets första levnadsår. Dock påpekar forskarna att studiens avsikt var att undersöka särskiljande karaktärsdrag, inte att utveckla ett screeninginstrument (Chang et al., 2021). Chang et al. (2021) menar att ögonskanning med datorsyn kommer ha stor betydelse för insamling av större dataset inom beteendeforskning med fokus på biomarkörer.

Imitation av rörelser. Vabalas et al. (2020) har valt att undersöka förmågan att imitera rörelser hos vuxna personer ($n = 44$, $m = 31.24$ år). Deltagarna delades två lika stora grupper med och utan AST, matchade på ålder, kön, huvudhand (höger/vänster) och intelligenskvot (IQ [Vabalas et al., 2020]). Deltagarna får se rörelsesekvenser på en skärm, där ett finger förflyttas mellan ett antal punkter. Deras uppgift är att själva utföra samma sekvens på bordet framför dem, med motsvarande punkter som i videon. Datainsamlingen gjordes med en rörelsedetektor och en ögonskanner, för att särskilja grupperna åt användes maskininläring.²⁸ Information från 30 deltagare användes som träningsmaterial och övriga 14 deltagare sparades till ett oberoende test. Antalet kännetecken som inkluderades begränsades till 10. Den högsta träffsäkerhet med klassificering av kroppsrörelser var 73%, sensitivitet 88% och specificitet 59%. För ögonrörelser var träffsäkerheten 70%, sensitivitet

²⁸ En Support Vector Machine (SVM) algoritm med en radical basis function (RBF) kernel används för klassificering av grupperna, men en 10-faldig korsvalidering. SVM är en vanlig metod som används vid linjära och icke-linjära klassificeringsproblem vid maskininläring (Kang et al., 2020a). En styrka med SVM är effektiv klassificering vid låga deltagarantal och att få parametrar för träning krävs (Washington et al., 2020).

43% och specificitet 97%. En kombination av kroppsrörelser och ögonrörelser höjde träffsäkerheten till 78%, sensitivitet 57% och specificitet 99% (Vabalas et al., 2020). För kroppsrörelser avgränsas resultatet till 6 utvalda kännetecken för rörelsernas acceleration och retardation, då de värdena visat på statistiska skillnader (Vabalas et al., 2020). Personer med autism var snabbare på att nå den högsta accelerationen in en rörelse ($t[42] = 2.5, p = .017, d = .75$) och den kraftigaste retardationen skedde senare i rörelsen ($t[29.4] = 3.0, p = .006, d = .9$ [Vabalas et al., 2020]). Personer utan AST hade reducerad acceleration och retardation samt högre vertikal amplitud (hur högt handen lyfts) när de skulle imitera ovanligt förhöjda rörelser. Personer med AST tenderar att bibehålla sitt vanliga beteendemönster och missade att imitera ovanliga rörelser. Det stämmer med tidigare studier där autister fokuserar på måluppfyllelse och är sämre på att imitera specifika stilar i ett rörelsemönster (Vabalas et al., 2020).

Virtuellvärld och ögonskanning. Visning av virtuell verklighet har gått framåt och idag behöver åskådaren inte längre bära speciella glasögon eller hjälmar för att ta till sig den virtuella världen. I stället kan effekten fås genom visning på 2D vägg, där skärmarna på sidorna är vinklade så att en inneslutande miljö skapas.²⁹ Alcañiz et al. (2021) har använt en ”virtuell värld” uppdelad i 14 sekvenser plus en nedräkning i början och slutet av upplevelsen. Den virtuella världen var formad som ett köpcentrum med butiker och underhållning, miljöer där visuell (bild), auditiv (ljud) och olfaktoriska (lukt) sinnen stimuleras. Försökspersonerna var 55 barn, 4-7 år, ($n = 35$ barn med AST och $n = 20$ utan AST) och de bar ögonskannandeglasögon (Alcañiz et al., 2021). Undersökning av uppmärksamhet för dynamiska föremål förekommer även i andra studier. Alcañiz et al. (2021) menar att barnets uppmärksamhet i en naturlig miljö där det förekommer rörliga objekt, bättre kan återspeglas i en ”virtuell värld” än vad undersökningar med rörlig bild på en skärm möjliggör.

Om barnen fokuserade på annat än det som var markerat som områden av intresse registrerades det som bakgrund. Område av intresse kunde vara en eller flera personer, där underkategorier var avatar (bild som representerar användaren)/annan karaktär, kropp/ansikte eller barn/vuxna. För föremål av intresse var underkategorier dynamiska (i rörelse) respektive statiska (stillastående). För att få en överblick över de olika kategorierna som används för att undersöka om det föreligger skillnader mellan barn eller utan AST. Av 144 fokusområden

²⁹ Tekniken forskarna har använt beskrivs som ”2D three-wall CAVE™”, där CAVE är förkortning för Cave Assisted Virtual Environment.

som var identifierade, visade 13 på signifikanta skillnader mellan grupperna. En begränsning hade satts till att högst 10 fokusområden fick inkluderas i en kategoriseringsmodell för att undvika överanpassning. Barn med AST iakttog fler föremål totalt och påtagligt fler statistiska objekt än kontrollgruppen. Specifika scener gav signifikanta skillnader på vem och var barnen tittade. (Alcañiz et al., 2021).

Det högsta resultat att urskilja grupperna uppnådde en träffsäkerhet på 86%, sensitivitet 91%, specificitet 82%, AUC .91, F1 poäng .87 och 9 inkluderade kännetecken (Alcañiz et al., 2021).³⁰ Ett flertal tidigare studier har ansett att barn utan AST ha ett större totalt iakttagande av socialt innehåll än barn med AST, speciellt om den totala mängden av stimuli är stort (Alcañiz et al., 2021). Alcañiz et al. (2021) får med sin virtuella miljö ett motsatt resultat. Barn med AST aktiverar mer visuell uppmärksamhet och åskådliggör ett totalt större iakttagande än barn utan AST. Det andra resultat var att barn med AST hade en signifikant högre preferens för att iaktta vuxna. I jämförelse om valet bestod i att iaktta ett barn eller en avatar, fanns en liten men signifikant skillnad, barn med AST föredrog att titta på barn i högre utsträckning än barn utan AST. Vidare fanns tre scener där barnen med AST såg signifikant mer på ansikten, både barn och vuxnas, än de utan AST. Barnen med AST har en högre preferens för kropp och ansikten (Alcañiz et al., 2021).

En översikt över deltagare i studier som använder maskininlärning för att särskilja individer med och utan AST redovisas i tabell 3. He et al. (2021) använder tre grupper i sin studie, högfungerande och lågfungerade AST, samt en kontrollgrupp utan AST, därför skiljer sig den studien i tabellens uppställning.

³⁰ Resultatet nåddes med en support vector machine algoritim.

Tabell 3

Översikt över deltagare i studier som använder maskininlärning för att särskilja individer med och utan AST.

| Författare | Inriktning på studien | Inkluderade | Antal deltagare | Med ASD | Medel ålder / SD | Antal flickor | Utan ASD | Medel ålder / SD | Antal flickor |
|--------------------------|--|--------------------------------------|-----------------|---------|------------------------|---------------|----------|----------------------|---------------|
| Alcañiz et al. (2021) | Ögonskanning, 3D virtuell värld | Barn, 4-7 år | 55 | 35 | 5.26 / 0.51 | - | 20 | 4.75 / 0.77 | - |
| Chang et al. (2021)* | Datorsyn, visning av video | Barn, 16-38 mån Inbjudna n = 1564 | 993* | 40 | - / - | - | 936 | 21.1 mån | 491 |
| He et al. (2021) | Datorsyn, ledtrådar om vart blicken ska riktas | Barn, förskola | 74 | 50 | - / - | 8 | 24 | 5.24 / 0.81 | 6 |
| | | Varav högfungerade AST | | 26 | 5.08 / 0.83 | 3 | - | - | - |
| | | Varav lågfungerade AST | | 34 | 4.98 / 1.09 | 5 | - | - | - |
| Kang et al. (2020a) | Ögonskanning, ansiktsprocessande | Barn, 3-6 år | 157 | 77 | 4.52 / 0.99 | 9 | 80 | 4.64 / 0.98 | 14 |
| Kang et al. (2020b) | EEG och ögonskanning | Barn, 3-6 år | 97 | 49 | 4.29 / 1.07 | 10 | 48 | 2.26 / 1.0 | 12 |
| Mazumdar et al. (2021)** | Ögonskanning, visning av stillbilder | Barn, 5-12 år | 28 | 14 | 8 år / - | - | 14 | 8 / - | - |
| Nag et al. (2020)*** | Ögonskanning visning av stillbilder | Barn, 6-17 år | 33 | 16 | 12.13 / - | 3 | 17 | 11.53 / - | 8 |
| Vabalas et al. (2021) | Ögonskanning och rörelsedetektor | Vuxna | 44 | 22 | 31.53 år / - | 7 | 22 | 30.9 / - | 7 |
| Wan et al. (2019) | Ögonskanning, visning av video | Barn 4-6 år | 74 | 37 | 4.6 år / 0.7 | 4 | 37 | 4.8 / 0.4 | 10 |
| Zhao et al. (2021) | Ansikte-ansikte samtal | Barn, 6-13 | 39 | 19 | 99.6 mån. / 108.8 mån. | 2 | 20 | 108 mån. / 27.0 mån. | 3 |

SD = standardavvikelse för ålder.

*17 deltagare med försenad utveckling eller försenat språk har inte ingått av någon i grupperna AST eller TD.

Könsfördelning är endast redovisad för samtliga deltagare.

**Studien har återanvänt ett dataset tidigare publicerat dataset utan att redovisa tydlig information om populationen.

*** Studien misslyckas med att nå ett signifikant resultat.

Studierna har redovisat olika mått på prestanda. En del uppger hur många kännetecken som har ingått i resultatet, men inte alla. En sammanställning av resultaten redovisas i tabell 4.

Tabell 4.*Redovisade mått av AST / inte AST med maskininlärning, plus inkluderande kännetecken.*

| Författare | Träff-säkerhet (accuracy) | Sensitivitet (true positive rate / recall) | Precision (positivt prediktions -värde) | Specificitet (true negative rate) | F1-värde | Area under the curve (AUC) | Antal använda kännetecken |
|---------------------------|---------------------------|--|---|-----------------------------------|----------|----------------------------|---------------------------|
| Alcañiz et al. (2021) | 86% | 91% | - | 82% | .87 | .91 | 9 |
| Chang et al. (2021)* | - | - | - | - | - | .90 | - |
| He et al. (2021)** | 93.24% | - | - | - | - | .938 | - |
| Kang et al. (2020a) | 84.17% | - | - | - | - | .89 | 120 |
| Kang et al. (2020b) | 85.44% | 85.44% | - | - | - | .93 | 32 |
| Mazumdar et al. (2021)*** | 59% | 68% | 57% | 50% | .61 | - | - |
| Nag et al. (2020)**** | - | - | - | - | - | - | - |
| Vabalas et al. (2021) | 78% | 57% | - | 99% | - | - | 10 |
| Wan et al. (2019) | 85.1% | 86.5% | - | 83.8% | - | - | - |
| Zhao et al. (2021) | 92.31% | 84.21% | - | 100% | - | .92 | 3 |

*Studien uttrycker att den har använt ett flertal kännetecken, men uppger inte hur många.

**I tabellen redovisas resultatet för en tvådelad uppdelning, AST/inte AST, antalet använda kännetecken framkommer inte. Resultatet är det bästa för maskininlärning. En indelning i tre grupper, högfungerande autism, lågfungerande autism och kontrollgrupp hade en träffsäkerhet 81.08% med 68 kännetecken.

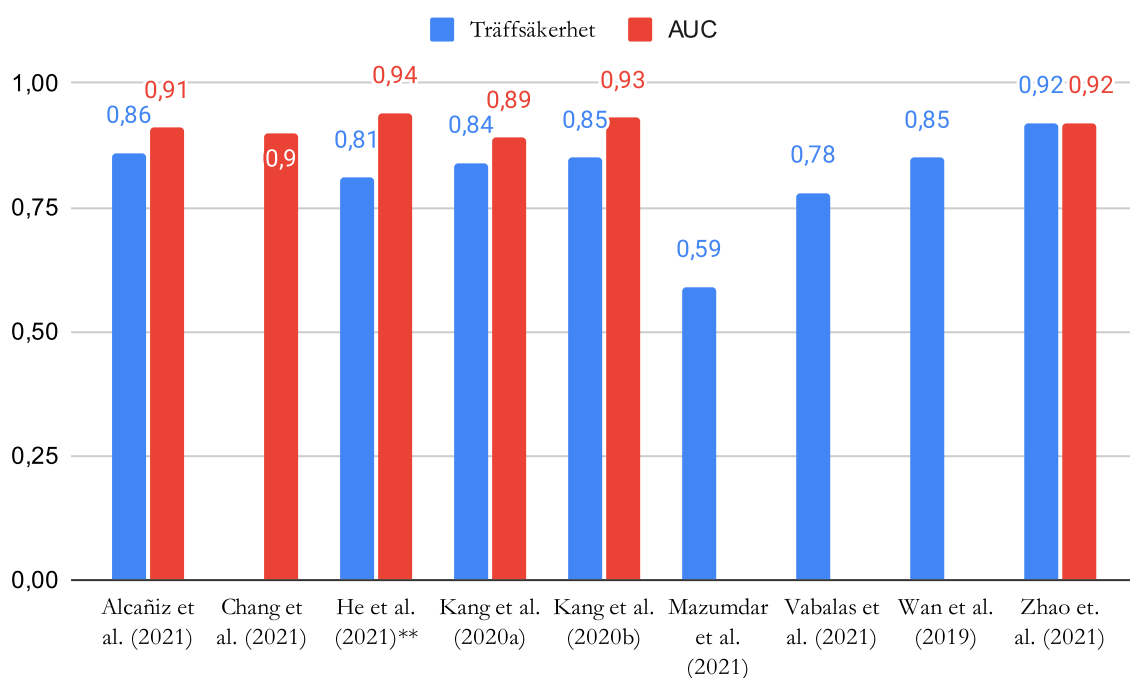
***Studien har återanvänt ett dataset tidigare publicerat dataset.

****Studien uppnår inga signifikanta resultat och har valt att inte redovisas några värden.

För att underlätta överblicken av resultatet redovisas träffsäkerheten och area under the curve (AUC) i ett stapeldiagram för maskininlärning. Nag et al.:s (2020) studie som ingick i maskininlärning, men som inte har redovisat något resultat då signifikans saknas har utelämnats. Se figur 7.

Figur 7.

Redovisade värden, träffsäkerhet och area under curve (AUC) för maskininlärning.



Fem studier har redovisat träffsäkerhet och AUC. Tre studier redovisar endast träffsäkerhet. Chang et al. (2021) är den enda som endast redovisar AUC. Träffsäkerhet har värdena spannet 0.59-0.92, för AUC är motsvarande 0.89-0.94.

Hur kan parvis jämförelse för ögonskanning förbättras? Idag dominerar

analyser av ögonrörelser som fokuserar på det spatiala (rumslig orientering). Den spatiala analysen fokuserar på vad blicken uppmärksammar mest och det som lyfts fram är skillnader i ”områden av intresse” (Król & Król, 2019b). Ett alternativ är temporal (tidsmässig) analys. Den intresserar sig för vad och i vilken ordningsföljd olika delar av en bild iakttas. Kunskapen om ordningsföljd kan användas för att bedöma sannolikheten i vad som iakttas när och om det finns tydliga preferenser för valet av ordningsföljd. Vad som kommer att iakttas först är, är i regel beroende av kontexten. Om en experimentgrupp som utsätts för påverkan från experiment ledaren jämförs med en kontrollgrupp där ingen påverkan sker, då är frågan stor andel av skillnaden beror på experimentet och vad skillnader mellan grupperna? Den problemställningen önskar Król och Król (2019b) att kunna komma runt genom att göra en parvis jämförelse.

För att testa skillnaden genomförs en studie med deltagare 44 matchade på ålder och intelligens, (n = 21 med AST 11-29 år, m = 16.27 år, varav 2 kvinnor och n = 23 utan AST 10-21 år, m = 16.31 år, varav 5 kvinnor [Król & Król, 2019b]). Om endast spatial

uppmärksamhet användes är resultatet $m = 53.9\%$. För spatial och temporal uppmärksamhet i kombination är resultatet $m = 55.5\%$. Skillnaden är signifikant $t(5) = -2.94$, $p = 0.032$ (Król & Król, 2019b). Men resultatet är sämre, då ett slumpmässigt resultat är 50%. Król och Król (2019b) försvarar sitt låga resultat med att det intressanta är den statistiska skillnaden mellan spatial uppmärksamhet jämfört med kombinationen av spatial och temporal uppmärksamhet. De menar att resultat sannolikt skulle förbättras om jämförelserna i tid var längre än de relativt korta 1,5 sekunders jämförelser som de har gjort (Król & Król, 2019b).

En översikt över deltagarna finns i tabell 5.

Tabell 5.

Översikt över deltagare i studien som använder maskininlärning med fokus på metodförbättring.

| Författare | Inriktning på studien | Inkluderade | Antal deltagare | Med ASD | Medel ålder / SD | Antal flickor | Utan ASD | Medel ålder / SD | Antal flickor |
|---------------------|--|-------------|-----------------|---------|------------------|---------------|----------|------------------|---------------|
| Król & Król (2019b) | Förbättra parvisa jämförelser med ögonskanning | 10-29 år | 44 | 21 | 16.27 / 4.84 | 2 | 23 | 16.31 / 2.69 | 5 |

SD = standardavvikelse för ålder.

Studier som urskiljer AST / inte AST med djupinlärning

Av de inkluderade studierna för djupinlärning har fyra av fem studier fokus på förskolebarn upp till tidig skolålder och en studie har inkluderat unga vuxna. De finns en stor variation över valet av undersökningsmetod eller kombinationer av flera metoder.

Avbildning av ögonens rörelsemönster. Forskarna Cilia et al. (2021) uttrycker en tydlig entusiasm för kombinationen spårning av ögonrörelser och maskininlärning vid diagnostik av neurologiska diagnoser. En illustration från studien visades när ögonskanning skulle förklaras (figur 1, sid 2). I studien inkluderades 59 barn, (29 med AST diagnos, $m = 7.7$ år, $SD = 2.6$ år, 10 flickor och 19 pojkar och 30 utan diagnos, $m = 8$ år, $SD = 2.8$ år, 11 flickor och 19 pojkar). Studien använder bild och videomaterial i syfte att testa barnens uppmärksamhet, medan deras ögonrörelser skannades. Ögonrörelsernas fokuspunkt registrerades och koordinater används för att se hur snabbt blicken rörde sig.³¹ Koordinaterna används för att visualisera ögonrörelserna i en svartvit kvadratisk bild. I bilden är varje sekvens av uppmärksamhet illustrerad som en vit linje mot en svart bakgrund. Den vita linjens kontraststyrka indikerar rörelsehastighet. Summan av linjer i bilden illustrerar blickens rörelsemönster inom en specifik tidsram. Bildens totalmönster används för att

³¹ De termer som Cilia et al. använder är point of gaze (POG) och the velocity of gaze movement (2021).

urskilja de som har autism och de som inte har autism genom en bildanalys med Convolutional Neural Network (CNN) algoritm, en form av djupinlärning. Resultatet visar en träffsäkerhet, $\approx 90\%$, sensitivitet, $\approx 83\%$, precision, $\approx 80\%$ och AUC .90 (Cilia et al., 2021). Själva sättet att omvandla ögonens rörelsemönster till en bild, samt bildernas användbarhet att effektivitet skilja mellan grupper är artikelns nytillkomna bidrag till forskningen (Cilia et al., 2021). Studien fann en stark korrelation mellan Childhood Autism Rating Scale (CARS) och hastigheten på ögonrörelserna, när metoden maximal informationskoefficient (MIC) används, (MIC = 0.79). Resultatet tolkas som en indikation för att bilder som visar blickens rörelsemönster kan användas för att bestämma graden av autism hos testpersoner (Cilia et al., 2021). Cilia et al. föreslår att en fördjupad studie skulle kunna inkludera fler bilder som handlar om annat än mellanmänsklig interaktion, då de med autism ofta har förhöjd uppmärksamhet på omgivningen (2021). Om uppmärksamhetstid är högre hos personer med autism, skulle det kunna fungera som en särskiljande faktor för diagnos.

Videoanalys. Li et al. (2020) har använt videofilmer för att försöka urskilja om 272 förskolebarn (136 med AST och 136 utan AST) från videofilmer (en film på ca 10 minuter per barn, mer precis ålder eller fördelning av kön framkom inte). En kombination av ett spårning-inlärning-upptäckande algoritm för videoanalys används efter att varje videospelning omvandlas till bildfrekvenser. Metoden använder varken ögonskanning eller datorsyn, utan spårar ögonrörelser genom analys av inspelat videomaterial. De sekvenser där ögonen är i bild registreras och rörelser lagras med ett koordinatsystem med en x- och y-axel. Identifierade sekvenser med ögonrörelser inkluderas i algoritmens inlärande del, så att analysförmågan förbättras med antalet granskade filmer. Sekvenser utan någon identifierad blick ges värdet noll och de delarna blir därmed bortsorterade. Barnen utan AST bibehåller oftare fokus på ett specifikt område. När deras ögonrörelser omvandlas till koordinater blir resultatet mer koncentrerat och ytan för blickens rörelseområde visar ett begränsat omfång. Barnen med AST uppvisar kortare uppmärksamhetsspann och fler ögonrörelser, deras samlade ögonrörelser i koordinatsystemet är spretigt och spritt på ett större område. För att dela in ögonens rörelsebanor i färre kategorier görs en fördelning i två övergripande kategorier. En kategori baseras på förändringsvinkeln mellan x- och y-koordinater och den andra kategorin är tiden som blicken är fokuserad på ett mindre område utan större rörelser. För varje kategori delas värdena in i åtta undergrupper för att underlätta för gruppvisa jämförelser av medelvärden. Analysen görs med djupinlärning. Resultat visar en träffsäkerhet 92.7%, sensitivitet 91.9% och specificitet 93.4% (Li et al., 2020). En slutsats i studien är att

djupinlärning kan öppna upp för att diagnostisering kan ske med inspelade filmer när barnet leker, pratar eller läser (Li et al., 2019).

Igenkänning av kroppshållning med video. Kojovic et al. (2021) har valt att fokusera på icke-verbala kännetecken och har inkluderat 136 barn, hälften med AST ($n = 68$, $m = 2.80$ år, $SD = 0.92$ år) och hälften utan AST ($n = 68$, $m = 2,55$ år, $SD = 0.97$ år). De har utgått från videomaterial där barnen har social interaktion genom fri lek med en vuxen forskningsledare, medan en anhörig till barnet finns i samma rum utan att delta i leken. Videons ljudupptagning har inte haft någon betydelse och allt ljudmaterial har exkluderats från när inspelningarna har analyserats. I videoanalysen filtreras bilden ett med dataprogram som kan visa skelettets uppskattade position som 2D bilder i realtid.³² Bilden med uppskattad kroppshållning påminner om enkla streckgubbar, där varje led är en markerad punkt. Interaktionen och rörelsemönster för de förenklade streckgubbarna analyseras med djupinlärning.³³ Därefter används 68 filmer för träning med maskininlärning för att urskilja grupperna åt. Resultatet baseras på 68 skilda filmsekvenser och redovisar F1-poäng .818, träffsäkerhet 80.9%, positivt prediktionsvärde 78.4%, sensitivitet 85.4% och specificitet 76.5% (Kojovic et al., 2021). Ytterligare ett test gjordes med 101 nya videofilmer för att testa modellens stabilitet, en träffsäkerhet 80.2% uppnåddes. Ett delresultat visade att en 70% träffsäkerhet nås med en 10-minuters video (Kojovic et al., 2021).

Fingerspårning. En alternativ teknik med samma syfte som ögonskanning är fingerspårning (Lio et al., 2020). För att använda fingerspårning behövs en skärm med tryckkänslighet, både smarta telefoner och surfplattor går bra. Metoden kräver ingen kalibrering. Den bild som visas på skärmen har en oskärpa som påminner om närsynthet. Om ett finger sätts på skärmen blir en cirkelformad yta precis ovanför fingret skarpt. Genom att föra fingret till ett område som väcker intresse, kan det utforskas på samma sätt som en blick som zoomar in på ett fokusområde. En jämförelse mellan ögonskanning och fingerspårning visar på en korrelation på .71.³⁴ Vidare gjordes en jämförelse om vilken metod som hade bäst förmåga att särskilja personer med AST ($n = 22$, $m = 20$ år, $SD = 2.5$ år, alla män) och en grupp utan AST ($n = 22$, matchade i ålder och kön).³⁵ För ögonskanning var AUC 0.862 och motsvarande för fingerspårning var AUC 0.908 (Lio et al., 2020). Lio menar att är

³² Dataprogrammet har namnet OpenPose (Kojovic et al., 2021).

³³ Här användes Convolutional Neural Network - Long short-term memory (CNN-LSTM) arkitektur.

³⁴ Två grupper fick granska 122 bilder, 61 bilder med ögonskanning och 61 bilder fingerspårning. Vilka bilder som visades med vilken metod och i vilken ordning, var motbalanserad mellan grupperna.

³⁵ Lio skriver ordagrant att kontrollgruppen är matchade efter ålder och kön, men alla inkluderade med AST är män. Det är oklart om Lio et al. (2020) har haft för avsikt att rekrytera kvinnor med AST, men misslyckats.

reliabiliteten för fingerspårning på pekskärm är hög och i paritet med ögonskanning (Lio et al., 2020).

EEG och ögonskanning. Zhang et al. (2021b) har valt att utvinna kännetecken från två skilda metoder, EEG och ögonskanning. Studien inkluderar 42 barn, (21 med AST och 21 utan AST, ålder 3-6 år). EEG undersökningen genomfördes först, därpå genomfördes deltester med igenkänning av egen eller främmande etnicitet, delad uppmärksamhet (joint attention) och ögonspårning. De två metodernas resultat har ingen tydlig korrelation med vardera. Maskininlärning har använts för att se vilka egenskaper som är mest särskiljande mellan barn med och utan AST för ögonskanning.³⁶ Resultaten av de två metoderna sammanfördes i en närhetsmatris med 14 listade egenskaper.³⁷ Egenskaperna för respektive värde i den nyskapade närhetsmatrisen undersöks med en Semi-Supervised Learning metod för graf-strukturerad data.³⁸ Metoden visar om noderna i närhetsmatrisen har koppling mellan varandra eller om de saknar koppling. Den erhållna informationen används för att urskilja vilka personer som har AST. Resultatet för analysen har en träffsäkerhet på 95.00%, precision 95.00%, sensitivitet 95.00% och F1-värde .9500 (Zhang et al., 2021b). När alla värden visar samma höga tal, väcks frågan om det är en takeffekt. Zhang et al. (2021b) kommenterar inte den frågan utan fokuserar på att den graf-strukturerade databehandlingen visar bättre träffsäkerhet än för renodlad djupinlärning.³⁹ Zhang et al. (2021b) tror att det beror på att resultatet med djupinlärning begränsas mer av ovidkommande brus i resultatbedömningen av mätningarna för EEG och ögonskanning.

En sammanställning av deltagare i studier som använder djupinlärning för att särskilja individer med och utan AST redovisas i tabell 6.

³⁶ En Random Forest algoritmen har använts för att undersöka de mest särskiljande egenskaperna i varje metod.

³⁷ Resultatet för särskiljning med EEG, träffsäkerhet 73.68%, precision 75.00%, sensitivitet 92.31% och F1-värde 0.8276. Resultatet för särskiljning med ögonskanning, träffsäkerhet 78.95%, precision 80.00%, sensitivitet 92.30% och F1-värde 0.8571.

³⁸ För att undersöka egenskaperna i den nyskapade närhetsmatrisen används en semi-supervised learning metod, Graph Convolutional Network, (GCN). En metod för att jämföra komplexa relationer i data med många egenskaper (Mayachita, 2020a).

³⁹ Vald metoden för djupinlärning var ResNet18. Redovisat resultat är träffsäkerhet 70.00%, precision 80.50%, sensitivitet 71.50% och F1-värde 0.6800.

Tabell 6.

Översikt över deltagare i studier som använder djupinlärning för att särskilja individer med och utan AST.

| Författare | Inriktning på studien | Inkluderade | Antal deltagare | Med ASD | Medel ålder / SD | Antal flickor | Utan ASD | Medel ålder / SD | Antal flickor |
|-----------------------|--------------------------------------|------------------|-----------------|---------|------------------|---------------|----------|------------------|---------------|
| Cilla et al. (2021) | Ögonskanning, stillbilder och videos | Barn, lågstadiet | 59 | 29 | 7.7 / 2.6 | 10 | 30 | 8 / 2.8 | 11 |
| Kojovic et al. (2021) | Datorsyn, kropps rörelser | Barn, 1.2-6.9 år | 136 | 68 | 2.80 / 0.92 | - | 68 | 2.55 / 0.97 | - |
| Li et al. (2020) | Datorsyn, videoinspelningar | Barn, 4-8 år | 272 | 136 | - / - | - | 136 | - / - | - |
| Lio et al. (2020) | Fingerskanning på skärm | Vuxna | 44 | 22 | 20 / 2.5 | 0 | 22 | - / - | 0 |
| Zhang et al. (2021b) | EEG och ögonskanning | Barn, 3-6 år | 42 | 21 | - / - | - | 21 | - / - | - |

SD = standardavvikelse för ålder.

En sammanställning över redovisade resultat i tabell 7, det skiljer sig mellan studierna om vilka kvalitetsmått som redovisas.

Tabell 7.

Redovisade mått för särskiljning av AST / inte AST med djupinlärning.

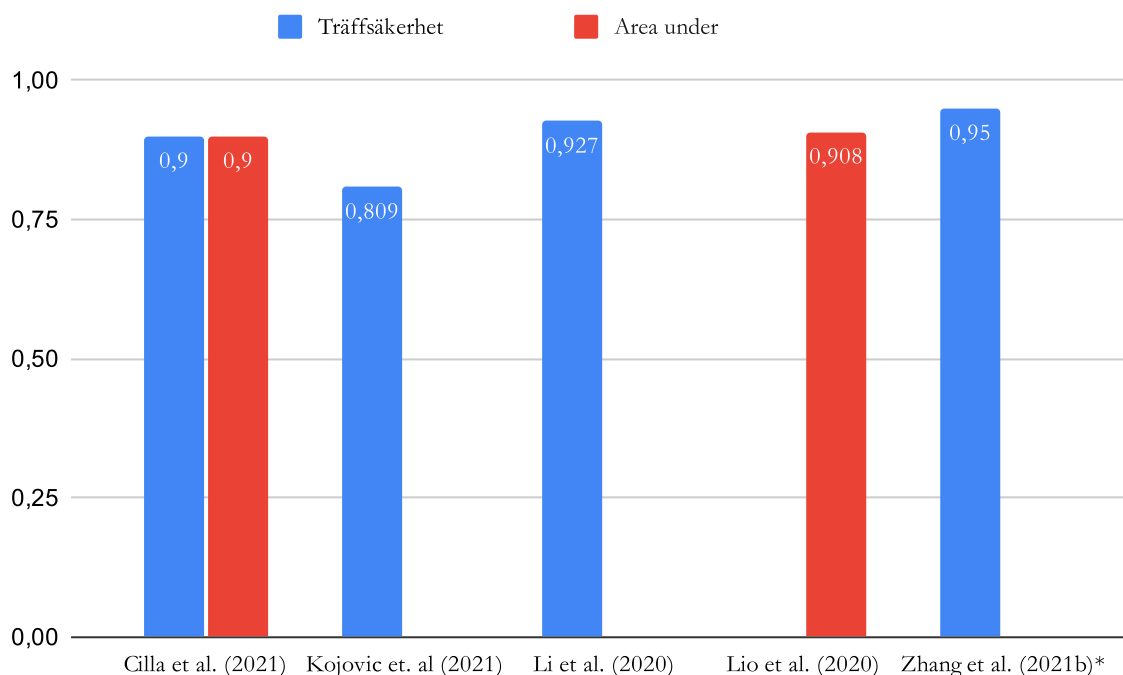
| Författare | Träff-säkerhet (accuracy) | Sensitivitet (true positive rate / recall) | Precision (positivt prediktions-värde) | Specificitet (true negative rate) | F1-värde | Area under the curve (AUC) |
|-----------------------|---------------------------|--|--|-----------------------------------|----------|----------------------------|
| Cilla et al. (2021) | 90% | 83% | 80% | - | - | .90 |
| Kojovic et al. (2021) | 80.9% | 85.4% | 78.4% | 76.5% | .818 | - |
| Li et al. (2020) | 92.7% | 91.9% | - | 93.4% | - | - |
| Lio et al. (2020) | - | - | - | - | - | .908 |
| Zhang et al. (2021b) | 95% | - | 95% | 95% | .95 | - |

SD = standardavvikelse för ålder.

Det som skiljer tabellen med resultat för djupinlärning från motsvarande tabell för maskininlärning är att frånvaron av antalet inkluderade kännetecken. För djupinlärning är det en okänd variabel, då processen från inmatning till resultat är dold. Träffsäkerheten och area under the curve (AUC) för djupinlärning visas i figur 8.

Figur 8.

Redovisade värden, träffsäkerhet och area under curve (AUC) för djupinlärning.



Anledningen till att det är glest mellan staplarna är att endast Cilla et al. (2021) har redovisat träffsäkerhet och AUC. De övriga studierna redovisat träffsäkerhet eller AUC. Träffsäkerhet har värden spannet 0.81-0.95, för AUC är motsvarande 0.90-0.93.

Diskussion

Vad krävs innan ögonskanningstekniker kan bli ett diagnosverktyg för AST inom psykiatrisk praktik? Det finns förväntningar på att AI i kombination med ögonskanning med ska kunna göra bedömningskriterier för AST mer objektiva, träffsäkrare och att testningen ska bli mer storskalig. Tilltro till en ökad skalbarhet grundas i att tekniken är relativt enkel och att undersökningar kan ske i vardagsmiljöer eller på mottagningar, vilket kan minska behovet av undersökningar på sjukhus. Samtidigt framstår övergången lång från forskningsstudier till allmänna kliniska verktyg. Nio praktiska problem som behöver överkommas vilka kommer att diskuteras här.

(1) Antalet inkluderade deltagare per studie behöver ökas. Studierna som med maskininlärning särskilde individer med AST och utan AST har ett medelvärde på 66.8 deltagare per studie, när Chang et al. (2021) exkluderats (n = 993).⁴⁰ Det kan jämföras med 11 studier med maskininlärning som ingick i Minissi systematiska litteraturstudie om AST med fokus på social visuell uppmärksamhet, 2015-2020. Tre studier är överlappande med de

⁴⁰ Król, M. & M. E., (2019b) studie för metodförbättringar har inte inkluderats i beräkningen (n = 44).

som är inkluderade här. Motsvarande medeltal för Minissi var 78.1 deltagare per studie. Dock var det endast 2 av 11 studier som hade ett deltagarantal som översteg 100 individer (Minissi et al., 2021). För djupinlärning var medeltalet 110.6 deltagare per studie.

När resultat från olika studier jämförs får varje procentenhet betydelse. Om antalet inkluderade deltagare är färre än hundra deltagare, så kommer varje deltagare att motsvara mer än 1-procentenhet per person. Enskilda personer vars resultat blir ofullständigt eller blir att betrakta som bortfall kan få en direkt inverkan på precisionen i resultatet. Allmänt är att ett ökat deltagarantal det enklaste sättet att öka styrkan mellan uppmätta statistiska samband (Shadish et al., 2002). För att slumpmässiga skillnader inte ska få ett betydande genomslag på resultatet är det önskvärt att deltagarantalet överstiger 100 personer, men helst inkluderar betydligt fler. Sju studier har mellan 28-60 deltagare (Alcañiz et al., 2021; Cilla et al., 2021; Lio et al., 2020, Nag et al., 2020; Vabalas et al., 2020; Zhang et al., 2021b och Zhao et al., 2021), se tabell 3 och 6. När det totala antalet deltagare visar upp till 60 kan studierna endast sägas ge en vägledning om vilket koncept som har bäst potential att testas i större skala. Men ge erfarenhet som bidrar till förbättrade experimentella upplägg och mätmetoder.

Reimann et al. (2020) visar att utmaningar bortom vad forskare kan påverka skapar problem med bortfall. Deras studie hade 50 inkluderade deltagare med AST, varav 17 föll bort, det motsvarar 1 av 3.

(2) Tillgång på dataset behöver förbättras. Det största hindret för utveckling av diagnostik med artificiell intelligens inom AST är tillgången till dataset. En möjlighet är att dataset delas. Redan idag kan dataset delas via öppna databaser som Kaggle.com och via forskningskonferenser. Men delade dataset kan medföra utmaningar, då de begränsar enskilda forskares inflytande över hur ett experiment har utformats, vilka deltagare som har rekryterats och vilka avvägningar som har gjorts. Det behöver finnas en tydlighet i att datainsamling och delning följer dataskyddsförordningens regelverk. Endast en forskargrupp har återanvänt ett delat dataset, Mazumdar et al, (2021) och studien var tillgänglig i ScienceDirect databas. I IEEE Explore,⁴¹ en databas som inte är inkluderad här, är det vanligt med publiceringar från forskningskonferenser. På konferenserna tävlar olika forskare om att uppnå högsta resultat utifrån samma delade dataset.

(3) Kodning av material behöver underlättas. Kodningens kvalitet påverkar

⁴¹ IEEE Xplore är en förkortning för Institute of Electrical and Electronics Engineers Xplore Digital Library.

resultatet och behöver ha god tillförlitlighet. Generellt är processer som sällar fram material som behöver extragranskas tidsbesparande, manuell granskning och kodning är ett tidskrävande arbete. För observationsstudier kan digital övervakning bespara experter tid. Vid observation av små barn kan datorsyn filtrera videoupptagningar till mer väsentligt material. Motsvarande filtrering gjordes i Li et al. (2020) studie för kroppsrörelser. Washington et al. (2019) förslog granskning av material kan fördelas till allmänheten genom webbtjänster. Graham et al. (2019) ser fördelar med att oövervakad maskininlärning minskar bias. Oavsett metodval för förenklad kodning vid maskininlärning behöver resultatet expertgranskas för att bedöma resultatets användbarhet och kvalitet, då träffsäkerhet, validitet och reliabilitet avgörs av träningsmaterialet. Vid djupinlärning kodas rådata automatiskt och det är mer oklart vad som har bedömts. Sannolikt är det därför Zhao et al. (2020) föreslår Semi-Supervised Learning, genom att det ger en bättre balans mellan automatisering, kvalitetskontroll och förståelse för hur materialet har processats.

(4) Normer för mätmetoder behöver etableras. När ett område är nytt finns det färre normer hur saker ska göras eller redovisas, i stället finns ett större fokus på att utforska och testa brett. Den här studien undersöker ett smalt område genom att fokusera på ögonskanning med analys av AI för att särskilja personer med AST från de utan AST. Antalet inkluderade studier är litet, men valet av metod visar stor variation.

Ett flertal studier har fokuserat på ”områden av intresse”, tillvägagångssätt varierar från stillbilder (Kang et al., 2020a; Kang et al., 2020b; Król & Król, 2019a; Król & Król, 2019b; Lio et al., 2020; Mazumdar et al., 2021) och videos (Chang et al., 2021; Li et al., 2020; Reimann et al., 2020; Sapiro et al., 2019; Wan et al., 2019) till virtuell verklighet (Alcañiz et al., 2021). Andra väljer att spåra blickens rörelsemönster (Cilia et al., 2021; He et al., 2021; Nag et al., 2020; Zhang et al., 2021b).

Typen av stimuli kan variera från förmågan att avläsa ansikten genom att bedöma känslouttryck (Król & Król, 2019a; Król & Król, 2019b; Nag et al., 2020), etnicitet (Kang et al., 2020a; Kang et al., 2020b; Zhang et al., 2021b) eller att härma rörelser (Vabalas et al., 2020). Andra studier har undersökt vad som väcker mest intresse sociala eller osociala stimulus (Alcañiz et al., 2021; Chang et al., 2021; Reimann et al., 2020; Sapiro et al., 2019; Wan et al., 2019), om individer med AST har högre fokus på centrumbias (Mazumdar et al., 2021), skillnader i spatial uppmärksamhet (Król & Król, 2019b) eller registrerat naturliga kroppsrörelser (Kojovic et al., 2021). Metoderna kommer att vidareutvecklas och vilken eller vilka som är bäst är långt ifrån avgjord.

Den enda skillnaden som framkommer mellan teknikerna ögonskanning och datorsyn är behovet av kalibrering (Zhao et al., 2021). Andra metoder för insamling av data som förekommit är EEG (Kang et al., 2020b; Zhang et al., 2021b), rörelsesensorer (Vabalas et al., 2020) och fingerskanning på pekskärm (Lio et al., 2020). Det finns skäl att tro att det dröjer innan någon form av standardisering tar plats eller att vissa metoder blir så pass dominanta att de uppfattas som norm. Om normer etableras, kan forskningen använda sina resurser mer samordnat och riktat.

Ibland är lösningarna förhållandevis enkla. Det Wan et al. (2019) uttryckte som ett problem, att det är svårt att få barn att titta på en skärm tillräckligt länge för att samla in relevanta mätvärden. Något som Chang et al. (2021) löste genom att låta dem sitta i knät på vårdnadshavaren och samla in information med datorsyn.

(5) Testresultat behöver förbättras och harmoniseras. Studierna visar på en stor spridning i träffsäkerhet (accuracy). För maskininlärning, 9 inkluderade studier (tabell 4) är träffsäkerhet i spannet 59-93% och för djupinlärning, 5 inkluderade studier 81-95% (tabell 7). Trenden är att enskilda studier för djupinlärning har högre träffsäkerhet än de för maskininlärning. Det är i linje med vad Quaak et al.:s (2021) kvantitativa metaanalys för neurologiska bilder med psykiatrisk inriktning kom fram till. De jämförde 29 studier med djupinlärning med 15 studier med maskininlärning (Quaak et al., 2021).

I den här litteraturstudien var den högsta uppvisade träffsäkerheten 95%. Om 95% träffsäkerhet skulle användas i klinisk verksamhet skulle 1 av 20 individer som väljer att utredas få ett felaktigt testresultat. Sett till en population blir det fort en stor grupp, därför behöver felmarginalen minska ytterligare. Resultatredovisningen i tabellerna 4 och 7 visar att det finns stora skillnader i vilka statistiska kvalitetsmått som redovisas. En harmonisering behövs för att tydliggöra hur olika metoder står sig mot varandra.

(6) Uppdelningen av ”AST” / ”inte AST” behöver nyanseras. Tre av studierna antyder att uppdelning i att ”ha” eller ”inte ha” AST blir onyanserat. Chang et al.:s (2021) studie med fokus på småbarn har en egen kategori för försenad verbal förmåga. Kang et al. (2020a) jämför lågfungerade autistiska barn med normalutvecklade barn. Kojovic et al. (2021) fokuserar på rörelsemönster och studien syftar till att utforma ett test som är oberoende av barns verbala förmåga. He et al.:s (2021) studie med ledtrådar om vart blicken ska riktas var den enda studie som jämfört resultatet från en tvådelad och en tredelad kategoriindelning. Vid en tvådelad uppdelning av förskolebarn med eller utan autism uppnås en träffsäkerhet på 93.24%. När antalet kategorier ökas till tre grupper, lågfungerade autism, högfungerande autism och normal utveckling, sjunker resultatets träffsäkerhet till 81.08%

(He et al., 2021). Det sänker resultatet med över 10 procentenheter och visar hur snabbt komplexiteten ökar när kategorier nyanseras. Den tredelade uppdelningen kan jämföras med Kang et al. vars resultat för barn med låg begåvning och en kontrollgrupp, där uppgiften var ansiktsgenkänning visade en träffsäkerhet 84.17% (och AUC .89 [Kang et al., 2020a]).

En idé om hur nyansering av AST kan testas kom från Cilia et al. (2021). De uttryckte att hastigheten på ögonrörelserna mät med metoden maximal informationskoefficient (MIC) skulle kunna vara en möjlig väg, men mer forskning behövs.

(7) Resultat behöver enkelt kunna förmedlas. Oavsett om de som ska ta del av ett utredningsresultat är barn, vuxen eller föräldrar så bör resultatet vara lättöverskådligt, begripligt och får gärna göras med en bildillustration. Av de inkluderade studierna så är visualiseringen av lagrade ögonrörelser som i Cilia et al. (2021) uppvisar ett exempel på (figur 1, sid 2), den enklaste framställningen som framkommit och resultatet redovisas på individnivå. I de resultatbilder Cilia et al. visar i sitt exempel har personer utan AST ett mer koncentrerat rörelsemönster, medan motsvarande mönster för personer med AST spretade betydligt mer. I kontakten med föräldrar får Cilia et al. bättre acceptans för resultatet för ögonskanning i relation till AST, än när gränsvärden på skalan för Autism Diagnostic Interview-Revised (ADI-R) används (2021).

He et al. (2021) försökte få barn att förstå med bildliga ledtrådar vart de skulle rikta blicken. De redovisar sitt resultat med mönstret av ögonrörelser placerat över experimentets bilder, vilket påminner om Cilia et al.:s (2021) resultatredovisning (figur 1, sid 2). Men He et al.:s (2021) sätt att redovisa kräver mer tankeprocessande för att tolkas.

Kang et al. (2020a) har i sin första studie markerat ”områden av intresse” med prickar. Där blicken har stannat länge skapas ett tätare kluster av prickar. I nästa studie har Kang et al. (2020b) övergått till en värmekarta, vilket även Mazumdar et al. (2020) använder för att redovisa sitt resultat. De områden som fått mer fokus får en varmare färg respektive intensivare ljus. De två studierna redovisar exempelbilder som jämför personer med och utan AST. Det kräver ett visst tankeprocessande för att tyda vad som skiljer dem åt. Bilderna är inte alls lika enkla att tolka som Cilia et al.:s (2021) resultatpresentation.

(8) Andelen flickor/kvinnor i studier behöver ökas. Få flickor eller ett skevt urval kan öka sannolikheten för överanpassning och försämrar den externa validiteten. Inom autismforskning finns ett befintligt mönster med en dominans av manliga deltagare (Mo et al., 2021). Av 17 inkluderade studier har 6 inte uppgivit könsfördelning för deltagarna (Alcañiz et al., 2021; Kojovic et al., 2021; Li et al., 2020; Mazumdar, 2021; Sapiro et al., 2019; och Zhang et al., 2021a). Två studier har endast rekryterat unga pojkar och män (Lio et

al., 2020, Reimann et al., 2020). He et al. (2021) uppger de har för få kvinnliga deltagare med autism för att undersöka om det finns genusrelaterade skillnader för social uppmärksamhet. Vabalas et al. (2021) lyfter fram att det var obalans mellan träningssetet med 20% kvinnor och den oberoende testningen med 57% kvinnor. Om det finns eventuella kännetecken som är typiska för flickor/kvinnor med AST och urvalet inte förändras, så kommer den variabeln sannolikt inte att bli tydligare med diagnostik utförd med AI.

(9) Kliniker behöver övertygas om teknikens fördelar och tillförlitlighet. När tekniker för att diagnostisera AST med ögonskanning eller datorsyn uppfyller krav på tillförlitlighet och tekniskmognad, återstår att övertyga kliniker. Det är en process som kan sammanfattas i begreppet ”diffusion of innovation”, vilket är ett eget forskningsfält med allmänna teorier om hur nya idéer och tekniker sprids. En indelning kan göras i fem kategorier efter turordning, innovatörer, tidiga anhängare, tidig majoritet, sen majoritet och efterslänrare (Dearing & Cox, 2017).⁴² De olika grupperna skiljer sig åt hur lätta de låter sig övertygas och vad attityden till det nya är. Om den nya tekniken uppfattas som ett hot mot den egna kompetensen, kan tekniken ses som fientlig och direkt motarbetas. Innan ögonskanning i kombination med AI för diagnostik av AST blir ett kliniskt verktyg, krävs flera år av ytterligare teknikutveckling och forskning. Därefter kan en utrullning av försöksverksamhet börja ta form. Det är svårt att förutse hur konkurrens med andra tekniker kommer att påverka utvecklingen, men sannolikt kommer nya verktyg och metoder att inkluderas i AST utredningar.

Begränsningar och svårigheter i litteraturgenomgången. En utökad sökning med ”machine learning + autism + computer vision” och ”deep learning + autism + computer vision” skulle vara möjlig, men har inte inkluderats. Inledningsvis var jag inte klar över att ögonskanning och datorsyn var så tydligt angränsande och överlappande metoder. För att begränsa studiens omfång blev valet att inte utöka antalet inkluderade artiklar med en utökad sökning. Om sökningen ”machine learning + autism + computer vision” hade inkluderats skulle 6 helt nya sökträffar tillkommit för närmare granskning. Sökningen ”deep learning” + autism + computer vision” skulle ge 4 helt nya sökträffar för närmare granskning. Ett par dubletter förekom, men de har räknats bort i summeringen. Det kan framstå som 10 artiklar är ett begränsat antal att lägga till, men det skulle kunna innebära att studiens omfång kunde ha ökat med ca. 50%.

⁴² Begreppet ”diffusion of innovation” och uppdelningen i fem grupper presenteras första gången 1962 i bok med samma titel skriven av E.M. Rogers, och (Dearing & Cox, 2017).

Studien med det i särklass högsta resultatet Zhang et al. (2021b) var en av de svårare att visuellt sätta sig in i. Zhang et al. förespråkar och använder Graph Convolution Network en grafisk illustrativ metod för att undersöka samband mellan två helt olika mätmetoder, EEG och ögonskanning. Mätresultaten omvandlas till grafer som läggs över varandra för att undersöka om det finns ett starkt eller svagt samband mellan graferna, genom att jämföra avstånd mellan punkter.⁴³ Ett kort avstånd mellan två punkter visar ett starkt samband och omvänt. Själva analysen sker i dolda lager vilket är ett kännetecken för djupinlärning (2021b). När tröskeln för att sätta sig in i resultatet är så pass högt, hur ska en testledare kunna förmedla resultatet vidare och kunna besvara frågor?

Begränsningar i granskningsprocessen. En svår avvägning i granskningsprocessen är att veta vad som är relevant att förmedla. För studenter som läser psykologi är terminologin, metodiken och skillnader mellan algoritmer inom maskin- och djupinlärning i regel ett obekant område. Stora litteraturstudier brukar följa principen att söka efter något smalt och specifikt. Det förutsätter att den som genomför en studie på förhand vet vilka avgränsningar som är rimliga. För den här studien skulle en möjlig begränsning vara att bara inkludera särskiljande studier. Studier som endast undersökte kännetecken var få och de har en längre väg att gå innan metoderna når praktisk klinisk nytta. En alternativ avgränsning är att endast undersöka maskininlärning eller djupinlärning. Andra möjliga avgränsningar är valet mellan rörliga och statiska stimulus bilder, eller att endast välja studier som använder minst två skilda undersökningsmetoder. Det kan vara valfri hjärnabbildningsmetod i kombination med ögonskanning, kroppsliga rörelsemönster eller något annat.⁴⁴ En smalare avgränsning skulle kunna ha gett utrymme till att inkludera sökningar med datorsyn eller att inkludera tekniska databaser.⁴⁵

Styrkor och svagheter i studien. Valet av den snäva tidsramen 2019-01-01 - 2021-11-30 kan ses som en styrka för att ämnet är aktuellt eller kan ses som en svaghet för att vara för smalt. Om tidsramen för sökta artiklar ökats med 2 år tillbaka i tid, ett tillägg för perioden 2016-01-01 - 2018-12-31, skulle antalet sökta artiklar för maskininlärning ökat med en studie och varit oförändrat för djupinlärning.

⁴³ Den som önskar sätta sig in i Graph Convolution Network föreslås börja med två metodbeskrivningar av Mayachita (2020a, 2020b).

⁴⁴ Tänkbara hjärnabbildningstekniker är elektroencefalografi (EEG), diffusions MR (dMRI), functional magnetic resonance imaging (fMRI), strukturell MR (sMRI), med flera.

⁴⁵ En teknisk databas som skulle kunna vara aktuell är IEEE Xplore. Där förekommer bland annat publiceringar från forskningskonferenser där deltagarna använder maskin- och djupinlärning på dataset som berör neuropsykiatriska sjukdomar.

En svaghet i studien är det begränsade tekniska djupet. Förklaringar vad som skiljer olika algoritmer från varandra, vad de är optimerade för och vad som är deras svagheter skulle kunna ha gjorts mer ingående. Omvänt är valet att på ett övergripande plan beskriva hur AI används utan att göra ämnet allt för tekniskt en styrka. En styrka eftersom målgruppen för studien är de som är intresserad av den tekniska utvecklingen inom psykologi, men inte behöver ha teknisk förståelse för programmering eller optimering av algoritmer.

Framtida forskning

Det största psykiatriska forskningsfältet med artificiell intelligens är depression. Komplexiteten vad gäller diagnosen depression speglas i bredden av de vitt skilda studier som finns i kombination med AI. Det finns studier som använder maskininlärning för att screena en viss målgrupp för depression i förebyggande syfte (Benfares, 2021; Zulfiker et al., 2021), andra granskar poster på sociala medier för att finna tecken eller indikationer på depression (Chiong et al., 2021), försök pågår för att personanpassa behandlingar vid svår depression (Sajjadian et al., 2021) och kartläggning av biomarkörer för depression fortskrider (Lin et al., 2021). I gränslandet lingvistik (allmän språkvetenskap), datavetenskap och AI finns området "Natural language Processing" (NLP). NLP gör att datorer kan tolka och analysera språk i talad eller skriven form. Det skapar nya möjligheter att bedöma måendet hos enskilda individer eller screena hela populationer genom att granska patientjournaler. En studie vid Lunds universitet av Kjell et al. (2021) bad människor att skriva egna bedömningar om sitt eget mående i en helt fritt generad text. Texterna analyserades med hjälp NLP och jämfördes med resultatet av skattningsskalor för att finna kännetecken som kan användas för att bedöma personens upplevda grad av depression och ångest, utifrån vad de skriver om sitt eget mående.

En litteraturöversikt som täcker in ett valfritt delområde av depression och artificiell intelligens skulle bidra med att sammanställa och tillgängliggöra kunskap om var forskningen är idag och vart den är på väg.

Slutsatser

Den första forskningsfrågan ”hur kan ögonskanning och angränsande tekniker används i kombination med maskininlärning för diagnostik av AST?” Den andra forskningsfrågan byter ut ordet ”maskininlärning” mot ”djupinlärning”. I utformningen av studier har de två teknikerna mycket gemensamt. Maskininlärning ger mer insyn i hur information värderas och processas än vad djupinlärning gör. Det har betydelse för att förstå vad som leder fram till resultatet. Tendensen är att djupinlärning är träffsäkrare än maskininlärning. Det högsta resultatet som framkom var en graf-strukturerad dataanalys med två skilda mätmetoder, ögonskanning och EEG, träffsäkerhet 95%, precisionen (positivt prediktionsvärde) 95% och specificiteten (tillförlitlighet) 95% (Zhang et al., 2021b).

Datainsamling med ögonskanning respektive datorsyn förekommer med maskin- och djupinlärning med fokus på statiska/rörliga bilder eller direkt mellanmänniskt bemötande. Det som undersökts för att särskilja om AST förekommer eller inte har varit: kroppsrörelser (Kojovic et al., 2021; Vabalas et al., 2020), avläsning av ansikten för känslouttryck (Król & Król, 2019a; Król & Król, 2019b; Nag et al., 2020) eller etnicitet (Kang et al., 2020a; Kang et al., 2020b; Zhang et al., 2021b), preferenser för sociala eller osociala stimulus (Alcañiz et al., 2021; Chang et al., 2021; Reimann et al., 2020; Sapiro et al., 2019; Wan et al., 2019), centrumbias (Mazumdar et al., 2021), spatial uppmärksamhet (Król & Król, 2019b) och ögonens rörelsemönster (Cilia et al., 2021; He et al., 2021; Li et al., 2020).

Två studier med djupinlärning saknade motsvarighet i studierna för maskininlärning, fingerskanning på pekskärm (Lio et al., 2020) som undersökte ”områden av intresse” och spårning av blickens rörelser från videoinspelningar (Li et al., 2020). Kang et al.:s (2020b) studie med maskininlärning och Zhang et al.:s (2021b) studie med graf-strukturerad dataanalys utgick från samma datainsamling, EEG och ögonskanning. Men analysmetoderna skiljer sig väsentligt. Kang et al. (2020b) har låtit en algoritm undersöka skillnader i resultat med en statistisk metod för oberoende t-test för respektive mätmetod (EEG och ögonskanning). Medan Zhang et al. (2021b) har gjort en direkt grafisk jämförelse, vilket är en betydligt mer komplex teknik.

Studiens syfte var att besvara frågan ”Vad som krävs innan ögonskanningstekniker i kombination med artificiell intelligens kan bli ett diagnosverktyg för AST inom psykiatrisk praktik? Frågan kan delas in i nio utmaningar som behöver övervinnas innan verktyg för AI teknik kan användas för diagnos av AST inom klinisk praktik. En fördjupning av varje utmaning utvecklades i diskussionsavsnittet.

Referenser

- Alcañiz, M., Chicchi-Giglioli, I. A., Carrasco-Ribelles, L. A., Marín-Morales, J., Minissi, M.E., Teruel-García, G., Sirera, M., & Abad, L. (2021). Eye gaze as a biomarker in the recognition of autism spectrum disorder using virtual reality and machine learning: A proof of concept for diagnosis. *Autism Research : Official Journal of the International Society for Autism Research*, 15(1), 131–145.
<https://doi.org/10.1002/aur.2636>
- Benfares, C., Akhrif, O., El Idrissi, Y. E. B., & Hamid, K. (2021). A clinical support system for classification and prediction of depression using machine learning methods. *Computational Intelligence*, 37(4), 1619–1632. doi: 10.1111/coin.12377
- Çalıyurt, O. (2021). Artificial Intelligence as a Psychiatric Diagnostic Tool Instead of DSM. *Alpha Psychiatry*, 22(3), A-XI-A-XIII. doi: 10.5152/alphapsychiatry.2021.101
- Chang, Z., Di Martino, J. M., Aiello, R., Baker, J., Carpenter, K., Compton, S., Davis, N., Eichner, B., Espinosa, S., Flowers, J., Franz, L., Harris, A., Howard, J., Perochon, S., Perrin, E. M., Krishnappa Babu, P. R., Spanos, M., Sullivan, C., Walter, B. K., & Kollins, S. H. (2021). Computational Methods to Measure Patterns of Gaze in Toddlers With Autism Spectrum Disorder. *JAMA Pediatrics*, 175(8), 827–836.
doi:10.1001/jamapediatrics.2021.0530
- Chiong, R., Budhi, G. S., Dhakal, S., & Chiong, F. (2021). A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, 135, 1-12.
<https://doi.org/10.1016/j.compbiomed.2021.104499>
- Cilia, F., Carette, R., Elbattah, M., Dequen, G., Guérin, J.-L., Bosche, J., Vandromme, L., & Le Driant, B. (2021). Computer-Aided Screening of Autism Spectrum Disorder: Eye-Tracking Study Using Data Visualization and Deep Learning. *JMIR Human Factors*, 8(4), 1-11. <https://humanfactors.jmir.org/2021/4/e27706>
- Crawford, K. (2021). *The atlas of AI. power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- de Saint Laurent, C. (2018). In Defence of Machine Learning: Debunking the Myths of Artificial Intelligence. *Europe's Journal of Psychology*, 14(4), 734–747.
<https://doi.org/10.5964/ejop.v14i4.1823>
- Dearing, J. W., & Cox, J. G. (2018). Diffusion Of Innovations Theory, Principles, And Practice. *Health Affairs*, 37(2), 183–190. <https://doi.org/10.1377/hlthaff.2017.1104>

- 1177 vårdguiden. (u.å.). *Autismspektrum tillstånd – AST*. Hämtad 1 mars 2022, från <https://www.1177.se/uppsala-lan/sjukdomar--besvar/hjarna-och-nerver/larande-forstaelse-och-minne/autism/>
- Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H.-C., & Jeste, D. V. (2019). Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Current psychiatry reports*, *21*(11), 1-18. <https://doi.org/10.1007/s11920-019-1094-0>
- He, Q., Wang, Q., Wu, Y., Yi, L., & Wei, K. (2021). Automatic classification of children with autism spectrum disorder by using a computerized visual-orienting task. *PsyCh Journal*, *10*(4), 550–565. <https://doi.org/10.1002/pchj.447>
- IBM. (u.å.). What is computer vision? Armonk, New York: IBM.
Hämtad 19 oktober, 2021, från <https://www.ibm.com/topics/computer-vision>
- Kang, J., Han, X., Hu, J.-F., Feng, H., & Li, X. (2020a). The study of the differences between low-functioning autistic children and typically developing children in the processing of the own-race and other-race faces by the machine learning approach. *Journal of Clinical Neuroscience*, *81*, 54–60.
<https://doi.org/10.1016/j.combiomed.2020.103722>
- Kang, J., Han, X., Song, J., Niu, Z., & Li, X. (2020b). The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. *Computers in Biology and Medicine*, *120*, 1-5.
<https://doi.org/10.1016/j.combiomed.2020.103722>
- Kjell, K., Johnsson, P., & Sikström, S. (2021). Freely Generated Word Responses Analyzed With Artificial Intelligence Predict Self-Reported Symptoms of Depression, Anxiety, and Worry. *Frontiers in Psychology*, *12*, 1–14.
- Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M. (2021). Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Scientific Reports*, *11*(1), 1–10.
<https://doi.org/10.1038/s41598-021-94378-z>
- Król, M. & M. E., (2019b). A Novel Eye Movement Data Transformation Technique that Preserves Temporal Information: A Demonstration in a Face Processing Task. *Sensors*, *19*(10), 2377, 1-11. <https://doi.org/10.3390/s19102377>
- Król, M. E., & Król, M. (2019a). A novel machine learning analysis of eye-tracking data reveals suboptimal visual information extraction from facial stimuli in individuals with autism. *Neuropsychologia*, *129*, 397–406.
<https://doi.org/10.1016/j.neuropsychologia.2019.04.022>

- Kumar P., Chauhan, R. Stephan, T. Shankar A., & Thakur, S. (2021). A Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection. *11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021*, 568-574.
doi: 10.1109/Confluence51648.2021.9377199
- Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H.-C., Paulus, M. P., Krystal, J. H., & Jeste, D. V. (2021). *Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom*, 9(6), 856-864. <https://doi.org/10.1016/j.bpsc.2021.02.001>
- Li, J., Zhong, Y., Han, J., Ouyang, G., Li, X., & Liu, H. (2020). Classifying ASD children with LSTM based on raw videos. *Neurocomputing*, 390, 226–238.
<https://doi.org/10.1016/j.neucom.2019.05.106>
- Lin, Z., Lawrence, W. R., Huang, Y., Lin, Q., & Gao, Y. (2021). Classifying depression using blood biomarkers: A large population study. *Journal of Psychiatric Research*, 140, 364–372. <https://doi.org/10.1016/j.jpsychires.2021.05.070>
- Lio, G., Fadda, R., Doneddu, G., Duhamel, J-R., & Sirigu, A. (2019). Digit-tracking as a new tactile interface for visual perception analysis. *Nature Communications*, 10(1), 1–13. <https://doi.org/10.1038/s41467-019-13285-0>
- Mayachita, I., (2020a) Getting the Intuition of Graph Neural Networks. *Analytics Vidhya*. Hämtad den 15 december, 2021 från <https://medium.com/analytics-vidhya/getting-the-intuition-of-graph-neural-networks-a30a2c34280d>
- Mayachita, I., (2020b) Understanding Graph Convolutional Networks for Node Classification. *Analytics Vidhya*. Hämtad den 15 december, 2021 från <https://towardsdatascience.com/understanding-graph-convolutional-networks-for-node-classification-a2bfdb7aba7b>
- Mazumdar, P., Arru, G., & Battisti, F. (2021). Early detection of children with Autism Spectrum Disorder based on visual exploration of images. *Signal Processing: Image Communication*, 94, 1-8. <https://doi.org/10.1016/j.image.2021.116184>
- Miller, C. H., Sacchet, M. D., & Gotlib, I. H. (2020). Support vector machines and affective science. *Emotion Review*, 12(4), 297–308. doi: 10.1177/1754073920930784
- Minissi, M. E., Chicchi Giglioli, I. A., Mantovani, F., & Alcañiz Raya, M. (2021). Assessment of the autism spectrum disorder based on machine learning and social visual attention: A systematic review. *Journal of Autism and Developmental Disorders*, 1-16. <https://doi.org/10.1007/s10803-021-05106-5>

- Misman, M. F., Samah, A. A., Ezudin, F. A., Majid, H. A., Shah, Z. A., Hashim, H., & Harun, M. F. (2019). Classification of Adults with Autism Spectrum Disorder using Deep Neural Network. *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS), Artificial Intelligence and Data Sciences (AiDAS), 2019 1st International Conference On*, 29–34.
doi: 10.1109/AiDAS47888.2019.8970823
- Mo, K., Sadoway, T., Bonato, S., Ameis, S. H., Anagnostou, E., Lerch, J. P., Taylor, M. J., & Lai, M.-C. (2021). Sex/gender differences in the human autistic brains: A systematic review of 20 years of neuroimaging research. *NeuroImage: Clinical*, *32*, 1-25.
<https://doi.org/10.1016/j.nicl.2021.102811>
- Nag, A., Haber, N., Voss, C., Winograd, T., Tamura, S., Daniels, J., Ma, J., Chiang, B., Ramachandran, S., Schwartz, J., Feinstein, C., & Wall, D. P. (2020). Toward continuous social phenotyping: Analyzing gaze patterns in an emotion recognition task for children with autism through wearable smart glasses. *Journal of Medical Internet Research*, *22*(4) 1-13. <https://doi.org/10.2196/13810>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Journal of Clinical Epidemiology*, *134*, 178–189. <https://doi.org/10.1136/bmj.n71>
- Quaak, M., van de Mortel, L., Thomas, R. M., & van Wingen, G. (2021). Deep learning applications for the classification of psychiatric disorders using neuroimaging data: Systematic review and meta-analysis. *NeuroImage: Clinical*, *30*, 1-21.
<https://doi.org/10.1016/j.nicl.2021.102584>
- Reimann, G. E., Walsh, C., Csumitta, K. D., McClure, P., Pereira, F., Martin, A., & Ramot, M. (2021). Gauging facial feature viewing preference as a stable individual trait in autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *14*(8), 1670–1683. <https://doi.org/10.1002/aur.2540>
- Sajjadian, M., Lam, R. W., Milev, R., Rotzinger, S., Frey, B. N., Soares, C. N., Parikh, S. V., Foster, J. A., Turecki, G., Müller, D. J., Strother, S. C., Farzan, F., Kennedy, S. H., & Uher, R. (2021). Machine learning in the prediction of depression treatment outcomes: A systematic review and meta-analysis. *Psychological Medicine*, *51*(16), 2742–2751. <https://doi.org/10.1017/S0033291721003871>

- Sapiro, G., Hashemi, J., & Dawson, G. (2019). Computer vision and behavioral phenotyping: an autism case study. *Current Opinion in Biomedical Engineering*, 9, 14–20.
<https://doi.org/10.1016/j.cobme.2018.12.002>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Statens beredning för medicinsk och social utvärdering. (2014). *Utvärdering av metoder i hälso- och sjukvården och insatser i socialtjänsten*. SBU:s handbok.
https://www.sbu.se/globalassets/ebm/metodbok/sbushandbok_bilaga10.pdf
- Tai, A. M. Y., Albuquerque, A., Carmona, N. E., Subramaniepillai, M., Cha, D. S., Sheko, M., Lee, Y., Mansur, R., & McIntyre, R. S. (2019). Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artificial Intelligence In Medicine*, 99, 1-11. <https://doi.org/10.1016/j.artmed.2019.101704>
- Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2020). Applying Machine Learning to Kinematic and Eye Movement Features of a Movement Imitation Task to Predict Autism Diagnosis. *Scientific Reports*, 10(1), 1–13.
<https://doi.org/10.1038/s41598-020-65384-4>
- Wan, G., Kong, X., Sun, B., Yu, S., Tu, Y., Park, J., Lang, C., Koh, M., Wei, Z., Feng, Z., Lin, Y., & Kong, J. (2019). Applying eye tracking to identify autism spectrum disorder in children. *Journal of Autism and Developmental Disorders*, 49(1), 209–215. <https://doi.org/10.1007/s10803-018-3690-y>
- Washington, P., Park, N., Srivastava, P., Voss, C., Kline, A., Varma, M., Tariq, Q., Kalantarian, H., Schwartz, J., Patnaik, R., Chrisman, B., Stockham, N., Paskov, K., Haber, N., & Wall, D. P. (2020). Data-Driven Diagnostics and the Potential of Mobile Artificial Intelligence for Digital Therapeutic Phenotyping in Computational Psychiatry. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, 5(8), 759–769. <https://doi.org/10.1016/j.bpsc.2019.11.015>
- Zhang, S., Chen D., Tang Y., & Zhang L. (2021b). Children ASD Evaluation Through Joint Analysis of EEG and Eye-Tracking Recordings with Graph Convolution Network. *Frontiers in Human Neuroscience*, 15, 1-10.
<https://doi.org/10.3389/fnhum.2021.651349>
- Zhang, Z., Li, G., Xu, Y., & Tang, X. (2021a). Application of Artificial Intelligence in the MRI Classification Task of Human Brain Neurological and Psychiatric Diseases: A Scoping Review. *Diagnostics (Basel, Switzerland)*, 11(8) 1-27.
<https://doi.org/10.3390/diagnostics11081402>

- Zhao, Z., Bao, Z., Zhang, Z., Deng, J., Cummins, N., Wang, H., Tao, J., & Schuller, B. (2020). Automatic Assessment of Depression From Speech via a Hierarchical Attention Transfer Network and Attention Autoencoders. *IEEE Journal of Selected Topics in Signal Processing, Selected Topics in Signal Processing, IEEE Journal of, IEEE J. Sel. Top. Signal Process*, 14(2), 423–434. doi 10.1109/JSTSP.2019.2955012
- Zhao, Z., Tang, H., Zhang, X., Qu, X., Hu, X., & Lu, J. (2021). Classification of Children With Autism and Typical Development Using Eye-Tracking Data From Face-to-Face Conversations: Machine Learning Model Development and Performance Evaluation. *Journal of Medical Internet Research*, 23(8), 1-14. <https://doi.org/10.2196/29328>
- Zulfiker, M. S., Kabir, N., Biswas, A. A., Nazneen, T., & Uddin, M. S. (2021). An in-depth analysis of machine learning approaches to predict depression. *Current Research in Behavioral Sciences*, 2, 1-23. <https://doi-org.ludwig.lub.lu.se/10.1016/j.crbeha.2021.100044>