

---

# Artificiell intelligens inom psykiatrisk diagnostik och vård

JONATAN PANULA

---

Artificiell intelligens (AI) har förutspått medföra stora förändringar inom så gott som alla delområden i samhället, men de stora förändringarna inom psykiatrisk diagnostik och vård har än så länge uteblivit. Psykiatrisk diagnostik stöder sig än idag på symtombaserad klassificering, och det är fortfarande oklart om de underliggande biologiska förändringarna följer samma gränsdragningar mellan olika syndrom. Detta kan vara en av orsakerna till att framtagningen av diagnostiska modeller varit långsammare inom psykiatrin än inom många andra specialiteter. Värdet av att använda neuroradiologiska eller neurofysiologiska undersökningsmodaliteter för att omklassificera patienter i symtombaserade diagnoser kan ifrågasättas, och prognostiska modeller för att förutspå relevanta utfall på individnivå kunde vara av större nytta för patienterna. Djupinlärning bygger på artificiella neurala nätverk. Den kan hitta mönster i data som är för komplexa för människan att greppa och kunde bana väg för individuell precisionsmedicin inom psykiatrin. Under de senaste åren har det gjorts enorma framsteg inom språkmodeller, och dessa kommer högst sannolikt att ha stor inverkan på psykiatrin, där diagnostik, riskkalkyl och även behandling i form av terapi till stor del är språkbaserade.

## Artificiell intelligens och maskininlärning

Med artificiell intelligens (AI) avses en programvara, vars mål är att imitera den mänskliga hjärnans förmåga att tänka. Artificiell generell intelligens innefattar ett än så länge hypotetiskt koncept om en programvara som, likt människan, kan tillämpa redan inlärd kunskap och lösa typer av problem den aldrig stött på. Alla AI-algoritmer som hittills utvecklats klassas som snäva AI-algoritmer, det vill säga de är utvecklade för att lösa ett specifikt problem och saknar förmågan att anpassa sig till problemlösning utanför detta område. En självkörande bil kan inte spela schack och en schackdator kan inte köra bil.

Maskininlärning (ML) omfattar de algoritmer inom AI som karakteriseras av en förmåga att lära sig själv genom observationer samt att tillämpa det inlärd på osedda data. ML delas grovt in i övervakad och oövervakad inlärning. Övervakade ML-algoritmer är självprogrammerande modeller, vars regler byggs upp av de mönster som observeras i indata-utdata-par (fråga-svar) som matas in av användaren. Modellen tränas på en del av datainnehållet och tillåts själv hitta de mönster som bäst ger det önskade utfallet. Efter det evalueras modellens precision genom klassificering av resterande osedda data. I kliniska applikationer (som exemplifierats i figur 1) består

### SKRIBENTEN

**Jonatan Panula**, ML, läkare under specialistutbildning inom neurologi, HUS Neurocentrum, doktorand vid Helsingfors universitet.

indata ofta av olika typer av patientdata (bilder, blodprov, frågeformulär, journaler e.d.), medan utdata utgörs av en kategorisk eller regressiv variabel av kliniskt intresse. Utfallet kan väljas från samma tidpunkt, vilket är fallet i en diagnostisk modell, medan prognostiska modeller kopplar ihop indata och ett framtida utfall som utdata. Diagnostiska modeller kan utvecklas i tvärsnittsstudier, där både indata och utdata kan samlas in vid samma tidpunkt, medan prognostiska modeller kräver antingen uppföljning tills de önskade utfallen inträffat eller retrospektivt insamlat indata, ifall patienterna undersöks för första gången när utfallen redan är kända. I oövervakad ML, som än så länge mer sällan applicerats på kliniska data, saknas de förutbestämda utfallen. Oövervakad ML används främst för att gruppera utgångsdata enligt observerade mönster (klustring) eller komprimera indata och skapa en representation bestående av färre variabler (dimensionsreduktion) (1).

---

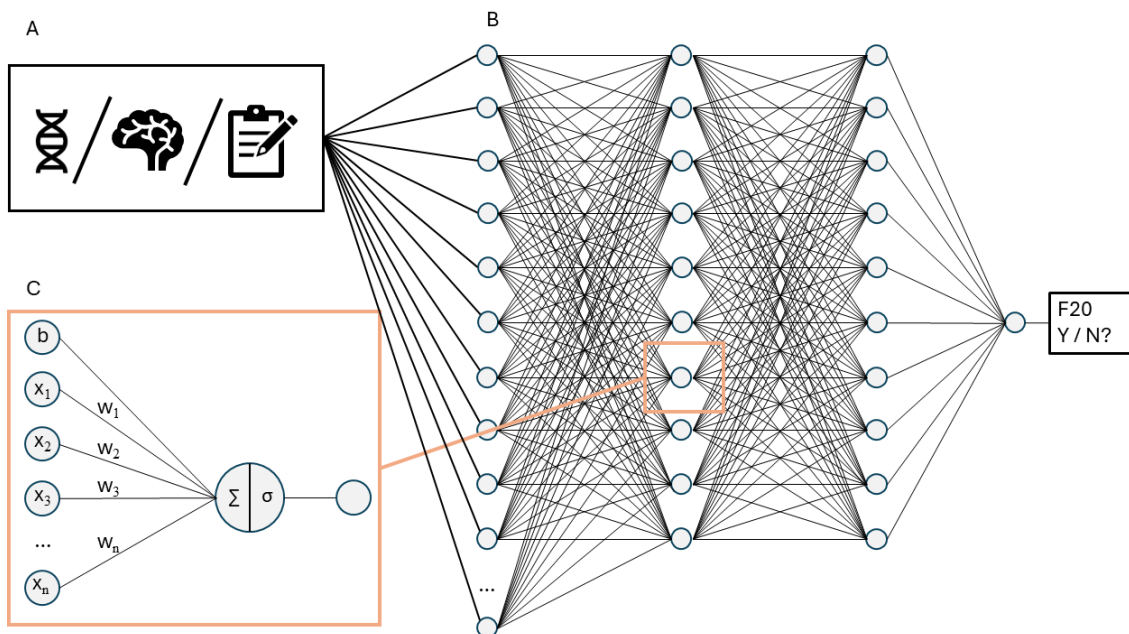
## Artificiella neurala nätverk och djupinlärning

Traditionella övervakade ML-algoritmer, som till exempel beslutsträd, logistisk regression och stödvektormaskiner, har använts i flera olika studier inom psykiatrin (2). Dessa algoritmers funktion blir lidande när mängden variabler ökar. För att tillämpa dessa algoritmer på data med upp till hundratusentals variabler (som text och bilddata) krävs ofta manipulation från användarens sida för att minska antalet variabler (dimensionsreduktion). Dessa problem har delvis lösts av djupinlärning (eng. *deep learning*), som bygger på komplexa artificiella neurala nätverk (ANN). Eftersom de modaliteter som teoretiskt sett kunde användas vid diagnostik inom psykiatrin (neuroradiologiska bilder, patientjournaler och genetiska analyser) innehåller ett stort antal variabler, anses djupinlärning vara en av de främsta kandidaterna för framtida diagnostiska och prediktiva modeller inom psykiatrin (3).

Den teoretiska grund som djupinlärning bygger på utvecklades för nästan ett sekel sedan. Intresset för och finansieringen av forskning inom artificiell intelligens har varierat, och under två så kallade AI-vintrar avtog optimismen och utvecklingen stannade upp så gott som helt (4). Under det senaste decenniet har intresset stadigt ökat, och de senaste åren har gett oss program som skapar konst, imiterar röster, kör bilar och skriver böcker. Dessa framsteg kan i första hand förknippas med ökad datorkraft, tillgång till massiva mängder data som modellerna tränas på samt nya ANN-algoritmer, som faltningsnätverk (eng. *convolutional neural network*) och transformatorer, som revolutionerat bild- respektive textanalys (5). ANN består av relativt enkla byggstenar, där den imponerande förmågan att lära sig komplexa samband uppkommer ur mängden noder och förbindelser mellan dem. De första grunderna till ANN härstammar från 1950-talet då perceptronen, en binär klassificerare, först utvecklades. Perceptronen skapades som en modell för hur neuroner antogs fungera och påminner till sin struktur om en mycket förenklad neuron. De artificiella neuroner som dagens djupa nätverk består av bygger fortfarande på samma principer. Likt perceptronen består den artificiella neuronerna av en eller flera inputnoder och en outputnod (figur 1). Varje inputnod får sin information från ett yttligare lager i nätverket, eller i fallet av det ytligaste lagret ur de data som

undersöks. Informationen i varje inputnod multipliceras med en vikt specifik för inputnoden i fråga, som representerar betydelsen eller styrkan av förbindelsen. Produkterna från varje inputnod (samt en bias) summeras i en aktiveringsfunktion, vars resultat förs vidare till nästa lager genom outputnoden. I perceptronen bestod aktiveringsfunktionen av en stegfunktion som imiterade aktionspotentialen, där binära utdata (0 eller 1) baserade sig på ett tröskelvärde som summan av inputnoderna jämfördes med. Moderna artificiella neuroner använder i stället icke-linjära aktiveringsfunktioner, vilket möjliggjort inlärning av mera komplexa samband mellan indata och utdata. De första perceptronnätverken bestod av ett hundratal perceptroner i två lager, medan de ANN som används idag kan bestå av miljarder artificiella neuroner i tiotals lager (6). Nätverket tränas med så kallade indatadata-par, det vill säga de data som matas in i kombination med det utfall som studeras. För ANN innebär detta att nätverket först analyserar indata, exempelvis så att styrkan i varje pixel i en bild utgör inputnoden för en neuron i indatalagret. I ett fullkopplat ANN kopplas varje neuron i ett yttligare lager till varje neuron i nästa lager, det vill säga varje neuron tar emot data från varje neuron i föregående lager. Utgångsvis är vikterna och bias slumpmässigt utvalda värden, och inga specifika aspekter av indata är betingade. Data passerar genom nätverkets neuroner och når slutligen den sista neuronerna eller de sista neuronerna i utdatalagret. Resultatet i utdatalagret jämförs med det kända utfallet för indata (exempelvis den korrekta diagnosen för den bild som visats) för att räkna ut en felvektor. Genom så kallad backpropagering modifieras vikterna och bias i nätverket lager för lager tillbaka genom nätverket för att stegvis föra det förutspådda utfallet närmare det korrekta. ANN tränas oftast upprepade gånger på tusentals indatadata-par. Följaktligen skapar nätverket regler för att ta fram mönster som karakteriserar korrekta utfall. För att undersöka hur väl nätverket fungerar analyseras slutligen data som nätverket inte tränats på, och modellens träffsäkerhet bedöms på basis av hur stor andel av de osedda fallen kategoriseras korrekt.

Av alla de framsteg som gjorts inom AI och djupinlärning under det senaste decenniet är de nya språkmodellerna (eng. *large language models*) kanske de mest omtalade. Teknikjättar har satsat enorma resurser på utveckling av sina egna algoritmer som GPT-4 från Open AI



**Figur 1.** Ett exempel på ett förenklat djupt neuralt nätverk för diagnosticering av schizofreni. A) Som indata kan olika typer av data användas, till exempel variabler ur genetiska undersökningar, strukturell eller funktionell magnetisk resonanstomografi, elektroencefalografi, neuropsykologiska undersökningsresultat eller patientjournaler. B) Det neurala nätverket består av ett indatalager, två dolda lager med tio noder samt ett utdatalager med en nod. De nätverk som används är ofta djupare och mer komplexa till sin struktur. C) I den artificiella neuronerna kombineras data från en eller flera inputnoder. Värdet från varje inputnod multipliceras med en vikt specifik för denna förbindelse. Produkterna från alla inputnoder och en bias summeras och körs efter detta igenom en aktiveringsfunktion, som modifierar värdet innan det slutligen presenteras som utdata.  $b$  = bias,  $w$  = vikt,  $x$  = en variabel i indata eller utdata från en nod i ett tidigare lager av nätverket,  $\Sigma$  = summan av bias och vikterna,  $\sigma$  = aktiveringsfunktion.

och Gemini från Google Deep Mind. Modellerna bygger på gigantiska AAN, även kallade transformatorer, som utvecklats för att kunna förutspå nästa ord i en sekvens och därmed skapa och tolka text. Företagens tillgång till datakraft och den massiva textmängd som finns tillgänglig på internet har möjliggjort utvecklingen av chattrobotar, som till skillnad från tidigare modeller har förmåga att förstå kontext, exempelvis sarkasm och humor, där samma ord kan ha olika värde och betydelse beroende på vilka andra ord det omges av.

Olika typer av ML-algoritmer kommer med hög sannolikhet att integreras i diagnostik och behandling de kommande åren. Ett flertal olika modeller med lovande resultat har redan utvecklats för diagnosticering av bland annat bröstcancer, multipel skleros, och retinopati (7).

## Maskininlärning inom psykiatri

### Diagnostiska modeller

Inom psykiatri kunde ML-baserade modeller medföra betydande förändringar. ML-algorit-

mer kunde erbjuda ett mycket välkommet hjälpmedel vid diagnostik, förutspå ogynnsamma utfall innan de inträffar, förbättra möjligheterna att hitta rätt farmakologisk eller neuromodulerande behandlingsform och även utgöra ett hjälpmedel i behandlingen genom integrering av språkmodeller.

Att klassificera psykiatriska patienter i de nuvarande deskriptiva diagnostiska kategorierna (DSM-V) baserat på olika typer av indata har varit ett aktivt forskningsområde under det senaste decenniet. Ett flertal relativt små studier har nått imponerande resultat vad gäller att skilja mellan patienter med schizofreni och friska kontrollpersoner (2, 8). Hög träffsäkerhet har nåtts med strukturell magnetisk resonanstomografi (MRT) (85 %) (9), enbaspolymorfi sammankopplad med funktionell MRT (87 %) (10) och elektroencefalografi (EEG) (99 %) (11). Mer atypiska former av indata, som tal (98 %) (12) och ansiktsuttryck (86 %) (13), har även gett goda resultat. Liknande resultat har uppnåtts för klassificering av depression, posttraumatiskt stressyndrom och bipolär sjukdom i små

---

kohorter (2, 14). I majoriteten av studierna tränades algoritmerna på patienter som hade insjuknat flera år tidigare. Eftersom medicinering eller sjukdom påverkar hjärnans struktur kan man inte utgå från att de fynd som görs flera år efter diagnos går att tillämpa på patienter som nyligen insjuknat. Pinaya et al. använde strukturella MRT-bilder och ett djupt ANN för att träna en modell som med 74 procents träffsäkerhet kunde klassificera patienter med schizofreni och kontrollpersoner, men när modellen tillämpades på patienter med en första psykos nådde den endast en träffsäkerhet på 56 procent (15). I de flesta studierna jämförs en grupp av patienter med en viss psykiatrisk diagnos med en kontrollgrupp bestående av psykiskt friska personer. Detta är fråga är sällan den som ställs till en kliniker. Om man exempelvis endast jämför patienter med depression och psykiskt friska personer har man inte byggt en metod för att diagnosticera depression, utan en modell för att observera avvikelser från ett normalt psykiskt tillstånd. Hur denna modell exempelvis klassificerar en bipolär sjukdom är omöjligt att förutspå. Schnack et al. använde strukturell MRT och en stödvektormaskin för att med 90 procents träffsäkerhet skilja mellan patienter med schizofreni och friska kontrollpersoner (16). Till skillnad från många andra studier tränade de även särskilda algoritmer för att skilja dessa två grupper från en tredje grupp bestående av patienter diagnosticerade med bipolär sjukdom, och kunde även då med 88 procents träffsäkerhet skilja mellan patienter med bipolär sjukdom och patienter med schizofreni. Trots att resultaten är lovande minskar studiens kliniska värde dock av att patienterna undersöktes i medeltal över tio år efter att de diagnosticerats. För att skapa en robust modell för diagnosticering måste kontrollgruppen innehålla olika psykiatriska komorbiditeter för att lära sig skilja dessa från det syndrom som undersöks. Detta framhävs av Taipale et al. som i sin studie visade att en modell som tränats att med god säkerhet klassificera patienter med schizofreni och friska kontrollpersoner felklassificerade nästan hälften av en grupp bestående av patienter med antisocial personlighetsstörning och substansmissbruk som schizofreni (17).

Psykiatriska diagnoser präglas av en markant heterogenitet och överlappning av symptom (18). Det är möjligt att diagnoser såsom schizofreni i verkligheten består av flera olika sjukdomar. Som exempel kan här nämnas behandlingsresistent schizofreni (där patienter

inte svarar på förstahandsbehandling med antipsykotika och ofta kräver behandling med klozapin), som möjligen har en annan biologisk bakgrund än schizofreni som svarar på den primära antipsykotiska behandlingen (19). Försök att hitta liknande, mera homogena underkategorier har gjorts genom att använda oövervakad maskininlärning. Till exempel använde Dwyer et al. strukturell MRT för att dela upp en kohort av patienter med schizofreni i två mindre grupper, där gruppen karakteriserades av distinkta degenerationsmönster och olika symtombild (20). För depression har liknande resultat uppnåtts. Drysdale et al. använde funktionell MRT för att kategorisera patienter med depression i fyra biotyper (21). Centralt för båda studierna var att träffsäkerheten för att skilja dessa biotyper från friska kontrollpersoner var klart bättre än när den ursprungliga patientkohorten jämfördes med kontrollpersoner.

### *Prognostiska modeller*

Prognostiska modeller och precisionsmedicin kunde vara av stort värde för psykiatrin, där sjukdomsförloppet kan vara svårt att förutspå och behandlingsresultaten ofta kan låta vänta på sig i flera veckor. Målet med precisionsmedicin är att i stället för att söka enskilda variabler som skiljer åt grupper hitta mönster som kan förutspå individuella utfall (22). Som Suvisaari et al. konstaterat är det fortfarande oklart om de kognitiva, genetiska, neurofysiologiska och neuroradiologiska variabler som korrelerar med olika kliniskt relevanta utfall i patienter med en första psykos på gruppnivå går att tillämpa på individnivå (23). Samma fråga kan ställas för övriga psykiatriska diagnoser. Nya resultat tyder dock på att ML-algoritmer, i alla fall delvis, kommer att kunna förutspå individuella utfall. Koutsouleris et al. kunde med 72 procents träffsäkerhet förutspå den framtida funktionsförmågan (bedömd med *global assessment of functioning scale*) hos patienter som ansågs löpa hög risk för att insjukna i en psykos (24), medan Leighton et al. med hög träffsäkerhet förutspådde arbetsförmågan i patienter med en första psykotisk episod ett år efter insjuknandet (25). En algoritm utvecklad av Chekroud et al. kunde med 65 procents träffsäkerhet förutspå om en patient med depression kommer att svara på behandling med citalopram (26). Vårdresistens och ett framtida behov av klozapinbehandling hos patienter med en första psykos är ett mycket relevant utfall sett i ett kliniskt

---

perspektiv. Osimo et al. använde kliniska variabler för att med 70 procents träffsäkerhet förutspå detta utfall (27), och vår forskningsgrupp publicerade nyligen en studie där vi genom att kombinera kliniska variabler med strukturella och funktionella MRT-bilder med 81 procents träffsäkerhet kunde förutspå vård-resistens (28).

Preliminära försök har även gjorts med att bygga modeller för att förutspå vilka patienter med hög risk för att insjukna i psykosjukdomar (eng. *at-risk mental states*, ARMS) som kommer att progrediera till en psykotisk episod. Till exempel använde Koutsouleris et al. en stödvektormaskin för att med 80 procents träffsäkerhet utgående från strukturella MRT-bilder förutspå vilka av de 73 undersökta ARMS-patienterna som skulle utveckla en psykotisk episod (29).

Det har, trots upprepade försök, visat sig vara mycket svårt att förutspå suicidförsök. Självdestruktivt beteende är högst sannolikt en komplex process som inte kan förklaras med enskilda variabler. Det är inte heller troligt att någon isolerad modalitet kan förklara alla enskilda fall. Många studier har nått träffsäkerheter över 70 procent genom att använda varierande typer av data. En nyligen publicerad översiktsartikel över ämnet (30) konstaterar att de studier där djupinlärning använts ger högre träffsäkerhet än traditionella ML-algoritmer. De flesta studierna använde endast en typ av indata och resultaten verifierades sällan på en extern kohort, vilket minskar studiernas värde sett ur ett kliniskt perspektiv. Ett återkommande problem i dessa studier är även den låga prevalensen av suicidförsök, vilket gör prospektiva studier svåra att genomföra. Den låga incidensen leder ofta till ett stort antal falska positiva resultat, vilket delvis kunde motverkas genom att bygga modeller för patientgrupper där incidensen är högre, och således uppnå ett högre positivt prediktivt värde.

### *Språkmodeller*

Med tanke på den centrala roll som språklig kommunikation har inom både psykiatrisk diagnostik och psykiatrisk vård är det föga förvånande att stora förväntningar ställts på språkmodeller. Under covid-19-pandemin skedde stora förändringar inom psykoterapi då en stor del av terapin genomfördes på distans. Enligt vissa studieresultat har behandlingssvaret varit varierande, men terapi på distans verkar vara en fungerande behand-

lingsform för en betydande del av patienterna (31), och många företag erbjuder nu enbart onlineterapi med goda resultat (32). Demzky et al. konstaterade, då de bad om stresshanteringsråd till en high school-studerande, att GPT-4 gav adekvata svar som till synes verkade relevanta och kunde tänkas komma från en studiehandedare, men inte stöds av nuvarande forskning (33). Modellen ger det mest sannolika svaret, men inte nödvändigtvis det svar som bäst stöds av forskningsresultat. Enligt Raile et al. ger programmen även bristfälliga och förenklade rekommendationer om behandling, när de konsulteras för second opinion angående vilken typ av terapi patienten kunde ha störst nytta av (34).

### **Framtidsutsikter**

Trots lovande resultat har ML-baserade modeller ännu inte medfört några stora förändringar inom psykiatrin (35). En del av de bakomliggande orsakerna är specifika för psykiatrin, medan andra är mera allmänna för integrering av ML i kliniskt arbete.

Till skillnad från den utförliga process som krävs när nya läkemedel införs saknas riktlinjer för hur och enligt vilka kriterier dessa program ska integreras i behandlingen (36). ML-algoritmer, och speciellt AAN, karakteriseras av en enorm komplexitet, vilket gör det svårt att förstå hur algoritmen kommer fram till rätt svar (37). De regler som klassificeringen bygger på uppstår i modellen ur själva data och förblir ofta dunkla för användaren. Reglerna begränsas av variationen i träningsdata, och algoritmen kommer att sträva efter att klassificera data baserat på de variabler som bäst möjliggör separering, oberoende om de har ett kausalt förhållande eller inte med exempelvis den sjukdom som undersöks. Likaså kan regler även uppstå ur oavsiktligt inkluderade variabler, exempelvis artefakter i radiologiska bilder. Att arbeta med verktyg, där varken läkaren eller patienten kan vara säker på varför ett behandlingsbeslut fattas leder till osäkerhet för båda parterna, och innan dessa frågor besvarats riskerar modellerna, trots häpnadsväckande resultat, förbli akademiska kuriositeter.

Det har visat sig vara väldigt svårt att bygga robusta övervakade ML-modeller för diagnostisk klassificering inom psykiatrin. En stor del av de studier som gjorts hittills bygger nämligen på små urvalsstorlekar, vilket ökar risken för överanpassning, där algoritmen blir skraddarsydd för de studerade patienterna

---

och inte går att generalisera till en större population (38). Träffsäkerheten tenderar att vara högre i små kohorter och kan förväntas sjunka när dessa modeller generaliseras till större och mera heterogena populationer (39). Till exempel visade Winterburn et al. i sin studie att resultatet blev sämre än väntat, när samma ML-modell användes på tre olika kohorter av patienter med schizofreni och friska kontrollpersoner (40). En orsak till att modeller som bygger på biologiska indata (neuroradiologiska undersökningar, elektrofysiologiska undersökningar och genetiska analyser) presterar suboptimalt kan ha att göra med annoteringen av data, vilket i diagnostiska modeller inom psykiatri är en deskriptiv diagnos. Trots att de nuvarande diagnostiska metoderna fungerar bättre än hittills presenterade alternativ för klinisk användning är det inte nödvändigt att de överensstämmer med underliggande biologiska förändringar. Om ML-modeller byggs på biologiska indata och deskriptiva symtombaserade diagnoser som utfall, kan modellens träffsäkerhet bli lidande ifall de symtombaserade skillnaderna inte har en klar biologisk korrelation som står att finna i just den modalitet som används som indata.

Små studier ökar även risken för diskriminering och rasism i modellerna (41). Socioekonomiska skillnader, tillgång till vård, rasism inom vården och varierande sjukdomsprevalens mellan etniciteter kan lätt oavsiktligt integreras i modellerna och försämra prognosen för enskilda patienter utan att det finns någon medicinsk grund för det. Kvinnor är ofta underrepresenterade i ML-studier, vilket leder till att modellen fungerar suboptimalt i de fall där kvinnors symptom skiljer sig från männens (41). Dessa problem kan motverkas genom större heterogena kohorter, men aktivt arbete krävs för att skapa modeller som inte reflekterar de orättvisor som redan finns inbyggda i samhället.

De modeller som byggs idag grundar sig ofta på en enda typ av indata, och ofta används bara en del av data för att förutspå utfall (dimensionsreduktion). Dimensionsreduktionen utförs antingen via a priori hypoteser om kausalitet eller AI-algoritmer som extraherar relevanta korrelationer ur träningsdata. Det är möjligt att de komplexa kognitiva funktioner som rubbas vid psykiatriska sjukdomar utgör så kallade lömska problem (eng. *wicked problem*) (42), där det problem som studeras inte kan undersökas i små delar eller isolerat från sin omgivning. Komplicerade fenomen som realitetsuppfattning och suicidalitet går

antagligen inte att reducera till ett tiotal variabler. Djupinlärning, som teoretiskt sett kunde bygga modeller baserat på strukturella eller funktionella MRT-bilder utan dimensionsreduktion, kunde vara ett möjligt redskap för att undersöka dessa fenomen. I de studier där djupinlärning jämförts med mer traditionella ML-metoder har djupinlärning ofta visat sig ge en bättre träffsäkerhet (8). Djupinlärning möjliggör även användning av flera modaliteter i samma modell, vilket antagligen kommer att behövas för att kunna förutspå komplicerade utfall som till exempel suicidförsök på individnivå.

Trots att kritiska röster har höjts angående de resultat som uppnåtts i små studier (38, 39) finns det ändå anledning till optimism. Språkmodellerna har gjort enorma framsteg under bara några år, och under det kommande decenniet får vi högst sannolikt se dem integrerade i terapi på distans. Språkmodeller kan redan nu imitera mänskligt tal, och det är sannolikt att en del av psykoterapin i framtiden utförs av en talande språkmodell, speciellt med hänsyn till den rådande bristen på psykoterapeuter. Dessutom har språkmodeller en häpnadsväckande förmåga att tolka text. Speciellt inom psykiatri, där majoriteten av diagnostik och riskkalkyl baserar sig på antingen text eller tal erbjuder språkmodeller ett attraktivt verktyg att integrera i behandlingen. En modell som omedelbart kan analysera alla journaler som skrivits om en patient samt de strukturerade intervjuer som kliniker (eller modellen själv) utför och jämföra dem med tusentals andra patienters journaler, där sjukdomsförloppet är känt kunde utgöra en av byggstenarna i ett kliniskt beslutsstödsystem. Det är inte otänkbart att en diskussion mellan en psykiater och dennes patient kunde bandas in och sammanfattas av en språkmodell som både patientjournal i patientdatasystemet och ett utkast till ett utlåtande som sedan godkänns av psykiatern. Integrering av språkmodeller för automatiserad textproduktion kunde avsevärt öka den tillgängliga tiden för patientinteraktion.

Prospektiva studier är mera tidskrävande och dyrare att utföra än tvärsnittsstudier, men prognostiska modeller är antagligen de modeller som kommer att ha störst inverkan på patienters välmående och klinikers sätt att jobba. De typer av data som ofta används i modellerna (såsom strukturell och funktionell MRT, EEG, genetiska analyser eller strukturerade intervjuer) är likaså arbetskrävande och dyra att framställa. Trots det bör ytterligare

försök göras att analysera flera olika typer av data i samma modell, vilket kommer att kräva djupare samarbete mellan forskningsgrupper. Sådana modeller kunde vara centrala för individanpassad medicin där potentialen inom psykiatri är enorm. Dessa prospektiva studier ska helst utföras i ett kritiskt skede, som till exempel när patienten behandlas för sin första psykos, som är en period när utfallen varierar stort och framtiden för patienten är svår att förutspå. Prognostiska modeller för patienter med en första psykos kunde leda till att direkt påbörja farmakologisk eller neuromodulerande behandling som med hög sannolikhet kommer att leda till symtomlindring i stället för att pröva sig fram, påbörja suicidpreventiva åtgärder för patienter med hög risk för suicidförsök eller snabbare sätta in rehabilitering för patienter som har goda förutsättningar att återgå till arbetslivet. Alla dessa åtgärder kunde märkbart förbättra vårdresultaten och ha en avsevärd inverkan på den enskilda patientens liv.

*Ett stort tack till min handledare, docenten Tuukka Raij, för inspirerande och konstruktiva kommentarer som bidragit till den slutliga texten.*

**Jonatan Panula**  
jonatan.panula@helsinki.fi

*Inga bindningar*

## Referenser

1. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput Sci.* 2021;2(3):160.
2. Iyortsuun NK, Kim SH, Jhon M, Yang HJ, Pant S. A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis. *Healthcare (Basel).* 2023;11(3).
3. Durstewitz D, Koppe G, Meyer-Lindenberg A. Deep neural networks in psychiatry. *Molecular Psychiatry.* 2019;24(11):1583–98.
4. Floridi L. AI and Its New Winter: from Myths to Realities. *Philosophy & Technology.* 2020;33(1):1-3.
5. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data.* 2021;8(1):53.
6. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med.* 2019;380(14):1347–58.
7. Shamshirband S, Fathi M, Dehzhangi A, Chronopoulos AT, Alinejad-Rokny H. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *J Biomed Inform.* 2021;113:103627.
8. Quaak M, van de Mortel L, Thomas RM, van Wingen G. Deep learning applications for the classification of psychiatric disorders using neuroimaging data: Systematic review and meta-analysis. *Neuroimage Clin.* 2021;30:102584.
9. Chen Z, Yan T, Wang E, Jiang H, Tang Y, Yu X, et al. Detecting Abnormal Brain Regions in Schizophrenia Using Structural MRI via Machine Learning. *Comput Intell Neurosci.* 2020;2020:6405930.
10. Yang H, Liu J, Sui J, Pearlson G, Calhoun VD. A Hybrid Machine Learning Method for Fusing fMRI and Genetic Data: Combining both Improves Classification of Schizophrenia. *Front Hum Neurosci.* 2010;4:192.
11. Sun J, Cao R, Zhou M, Hussain W, Wang B, Xue J, et al. A hybrid deep neural network for classification of schizophrenia using EEG Data. *Sci Rep.* 2021;11(1):4706.
12. Fu J, Yang S, He F, He L, Li Y, Zhang J, et al. Sch-net: a deep learning architecture for automatic detection of schizophrenia. *Biomed Eng Online.* 2021;20(1):75.
13. Huang J, Zhao Y, Qu W, Tian Z, Tan Y, Wang Z, et al. Automatic recognition of schizophrenia from facial videos using 3D convolutional neural network. *Asian J Psychiatr.* 2022;77:103263.
14. Chung J, Teo J. Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. *Applied Computational Intelligence and Soft Computing.* 2022;2022:1–19.
15. Pinaya WH, Gadelha A, Doyle OM, Noto C, Zugman A, Cordeiro Q, et al. Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. *Sci Rep.* 2016;6:38897.
16. Schnack HG, Nieuwenhuis M, van Haren NE, Abramovic L, Scheewe TW, Brouwer RM, et al. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage.* 2014;84:299–306.
17. Taipale M, Tiihonen J, Korhonen J, Popovic D, Vaurio O, Lahteenvuo M, et al. Effects of Substance Use and Antisocial Personality on Neuroimaging-Based Machine Learning Prediction of Schizophrenia. *Schizophr Bull.* 2023;49(6):1568–78.
18. Allsopp K, Read J, Corcoran R, Kinderman P. Heterogeneity in psychiatric diagnostic classification. *Psychiatry Res.* 2019;279:15–22.
19. Potkin SG, Kane JM, Correll CU, Lindenmayer JP, Agid O, Marder SR, et al. The neurobiology of treatment-resistant schizophrenia: paths to antipsychotic resistance and a roadmap for future research. *NPJ Schizophr.* 2020;6(1):1.
20. Dwyer DB, Cabral C, Kambeitz-Ilankovic L, Sanfelici R, Kambeitz J, Calhoun V, et al. Brain Subtyping Enhances The Neuroanatomical Discrimination of Schizophrenia. *Schizophr Bull.* 2018;44(5):1060–9.
21. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med.* 2017;23(1):28–38.
22. Dhamala E, Yeo BTT, Holmes AJ. One Size Does Not Fit All: Methodological Considerations for Brain-Based Predictive Modeling in Psychiatry. *Biol Psychiatry.* 2023;93(8):717–28.
23. Suvisaari J, Mantere O, Keinänen J, Mäntylä T, Rikandi E, Lindgren M, et al. Is It Possible to Predict the Future in First-Episode Psychosis? *Frontiers in Psychiatry.* 2018;9.
24. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, Rosen M, Rief A, Dwyer DB, et al. Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or With Recent-Onset Depression: A Multimodal, Multisite Machine Learning Analysis. *JAMA Psychiatry.* 2018;75(11):1156–72.
25. Leighton SP, Upthegrove R, Krishnadas R, Benros ME, Broome MR, Gkoutos GV, et al. Development and validation of multivariable prediction models of remission, recovery, and quality of life outcomes in people with first episode psychosis: a machine learning approach. *The Lancet Digital Health.* 2019;1(6):e261-e70.
26. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry.* 2016;3(3):243–50.
27. Osimo EF, Perry BI, Mallikarjun P, Pritchard M, Lewis J, Kautunda A, et al. Predicting treatment resistance from first-episode psychosis using routinely collected clinical information. *Nat Ment Health.* 2023;1(1):25–35.
28. Panula JM, Gotsopoulos A, Alho J, Suvisaari J, Lindgren M, Kiesepää T, et al. Multimodal prediction of the need of clozapine in treatment resistant schizophrenia; a pilot study in first-episode psychosis. *Biomarkers in Neuropsychiatry.* 2024;11:100102.
29. Koutsouleris N, Riecher-Rössler A, Meisenzahl EM, Smieskova R, Studerus E, Kambeitz-Ilankovic L, et al. Detecting the psychosis prodrome across high-risk populations using neuroanatomical biomarkers. *Schizophr Bull.* 2015;41(2):471–82.
30. Pignoni A, Delvecchio G, Turtulici N, Madonna D, Pietrini P, Cecchetti L, et al. Machine learning and the prediction of suicide in psychiatric populations: a systematic review. *Transl Psychiatry.* 2024;14(1):140.
31. Giordano C, Ambrosiano I, Graffeo MT, Di Caro A, Gullo S. The transition to online psychotherapy during the pandemic: a qualitative study on patients' perspectives. *Res Psychother.* 2022;25(3).

- 
32. Marcelle ET, Nolting L, Hinshaw SP, Aguilera A. Effectiveness of a Multimodal Digital Psychotherapy Platform for Adult Depression: A Naturalistic Feasibility Study. *JMIR Mhealth Uhealth*. 2019;7(1):e10948.
  33. Demszky D, Yang D, Yeager DS, Bryan CJ, Clapper M, Chandhok S, et al. Using large language models in psychology. *Nature Reviews Psychology*. 2023;2(11):688–701.
  34. Raile P. The usefulness of ChatGPT for psychotherapists and patients. *Humanities and Social Sciences Communications*. 2024;11(1):47.
  35. Meehan AJ, Lewis SJ, Fazel S, Fusar-Poli P, Steyerberg EW, Stahl D, et al. Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. *Molecular Psychiatry*. 2022;27(6):2700–8.
  36. van de Sande D, Van Genderen ME, Smit JM, Huiskens J, Visser JJ, Veen RER, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. *BMJ Health Care Inform*. 2022;29(1).
  37. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3:17.
  38. Chekroud AM, Hawrilenko M, Loho H, Bondar J, Gueorguieva R, Hasan A, et al. Illusory generalizability of clinical prediction models. *Science*. 2024;383(6679):164–7.
  39. Vieira S, Gong QY, Pinaya WHL, Scarpazza C, Tognin S, Crespo-Facorro B, et al. Using Machine Learning and Structural Neuroimaging to Detect First Episode Psychosis: Reconsidering the Evidence. *Schizophr Bull*. 2020;46(1):17–26.
  40. Winterburn JL, Voineskos AN, Devenyi GA, Plitman E, de la Fuente-Sandoval C, Bhagwat N, et al. Can we accurately classify schizophrenia patients from healthy controls using magnetic resonance imaging and machine learning? A multi-method and multi-dataset study. *Schizophrenia Research*. 2019;214:3–10.
  41. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical Machine Learning in Healthcare. *Annu Rev Biomed Data Sci*. 2021;4:123–44.
  42. Grisogono AM. How Could Future AI Help Tackle Global Complex Problems? *Front Robot AI*. 2020;7:50.

---

## Summary

### ***Artificial intelligence in psychiatric diagnosis and treatment***

*New treatment options and diagnostic models are eagerly awaited in psychiatry. Artificial intelligence is advancing ever more rapidly, but major impacts in clinical practice, especially regarding psychiatry, have yet to materialize. Psychiatric diagnoses are still symptom-based, and no biomarkers are routinely used in diagnostics. This may be a challenge for machine learning models, which to date rely largely on biomarkers. Prognostic models and personalized medicine could be of more value for patients and clinicians compared to diagnostic models. Due to the complex nature of psychiatric disorders, deep learning could provide a superior method for personalized medicine in psychiatry compared to traditional machine learning models.*