

Kunstmatige intelligentie voor klinische toepassing van beeldvorming in de psychiatrie

Guido van Wingen

Prof. dr. Guido van Wingen

Amsterdam UMC, locatie AMC

Universiteit van Amsterdam

Afdeling Psychiatrie

Amsterdam Neuroscience

Meibergdreef 5

1100 DD Amsterdam

E-mail: g.a.vanwingen@amsterdamumc.nl

Keywords: artificial intelligence; machine learning; neuroimaging; MRI; Psychiatry

Aantal woorden:

Samenvatting: 137

Hoofdstekst: 1579 / 1600

Aantal referenties: 20

Samenvatting

De klinische toepassing van beeldvorming bij psychische klachten is tot op heden gelimiteerd tot het uitsluiten van somatische pathologie. Radiologische beoordeling van hersenscans biedt doorgaans geen verklaring voor de psychische symptomen. Dat betekent echter niet dat psychische symptomen geen neurobiologische basis hebben. De hoop is daarom al jaren gevestigd op functionele MRI, waarmee de activiteit van de hersenen wordt gemeten. Dat heeft echter nog niet geresulteerd in klinische toepassingen. Een multivariate benadering middels machine learning analyse lijkt daar nu verandering in te gaan brengen. Machine learning algoritmes kunnen al verschillende taken in de radiologie automatiseren.¹ Recente studies laten zien dat machine learning analyse van MRI beelden ook diagnostische, prognostische en voorspellende biomarkers voor de psychiatrie op kan leveren. Grotere studies zijn nodig om klinische toepassingen te ontwikkelen, zoals clinical decision support systems ter ondersteuning van gepersonaliseerde behandelkeuzes.

Abstract

The clinical application of neuroimaging for psychological complaints has so far been limited to the exclusion of somatic pathology. Radiological assessment of brain scans usually does not explain the psychological symptoms. However, that does not mean that psychological symptoms have no neurobiological basis. Hope has therefore been placed on functional MRI, which measures the activity of the brain. However, this has not yet resulted in clinical applications. A multivariate approach using machine learning analysis now appears to be changing this. Machine learning algorithms can already automate various tasks in radiology. Recent studies show that machine learning analysis of MRI images can also provide diagnostic, prognostic and predictive biomarkers for psychiatry. Larger studies are needed to develop clinical applications, such as clinical decision support systems to support personalized treatment choices.

Beeldvorming in de psychiatrie

Beeldvormend onderzoek heeft de afgelopen decennia duidelijk gemaakt dat psychiatrische stoornissen gepaard gaan met structurele en functionele hersenafwijkingen. Wereldwijde samenwerking binnen het ENIGMA consortium (<http://enigma.ini.usc.edu/>) heeft geleid tot studies met duizenden patiënten. De unieke benadering van ENIGMA waarbij meta-analyses worden uitgevoerd op data van alle deelnemende onderzoekscentra maakt het mogelijk om representatieve effectgroottes te berekenen. De structurele afwijkingen die op deze manier zijn gevonden zijn echter dermate klein dat deze geen klinische rol van betekenis spelen. Zo is een kleinere hippocampus een van de meest consistente bevindingen, met slechts kleine tot middelgrote effecten (Cohen's $d < 0.5$).² De hippocampus vertoont daarom meer overeenkomst dan verschil ten opzichte van gezonde mensen. Ter vergelijking, de American Psychiatric Association stelt dat biomarkers een onderscheidend vermogen dienen te hebben van 80%.³ Dat is vergelijkbaar met een Cohen's d van 1.53 en daarmee een orde van grootte verschil. Vergelijkbare analyse van functionele MRI data is op dit moment nog niet beschikbaar, maar meta-analyse van gepubliceerde resultaten laat consistente hypoactivatie van de prefrontale cortex zien en hyperresponsiviteit van de amygdala en hippocampus op emotionele stimuli.^{4,5} Voor zulke studies kunnen echter geen effectgroottes worden berekend, maar ook die zullen vermoedelijk niet groot genoeg zijn voor klinische toepassing. Dit type MRI onderzoek is erop gericht om specifieke hersenafwijkingen te lokaliseren. Maar het is de afgelopen jaren juist duidelijk geworden dat psychiatrische symptomen vooral gerelateerd zijn aan verstoorde communicatie tussen meerdere hersengebieden.

Machine learning

Waar reguliere univariate MRI statistiek de informatie van elk hersengebied (of 'voxel', een volume pixel) onafhankelijk analyseert, maakt multivariate machine learning analyse het mogelijk om de hersenen als geheel te benaderen. Een ander belangrijk verschil is de doelstelling: waar univariate statistiek uitspraken doet over de gemiddelde patiënt, maken machine learning modellen voorspellingen voor individuele patiënten. Machine learning modellen dienen echter wel altijd gevalideerd te worden met onafhankelijke data. Dat gebeurt doorgaans middels kruisvalidatie. Daarbij wordt een deel van de data opzij gezet tijdens het maken ('trainen') van modellen, wat vervolgens wordt gebruikt om het model te testen. Dit proces wordt vervolgens herhaald voor een ander deel van de data en levert een uitkomst op die aangeeft hoe goed een model voorspellingen maakt voor nieuwe doch vergelijkbare data. Maar aangezien data flink kunnen variëren van ziekenhuis tot ziekenhuis, bijvoorbeeld door verschillen in de lokale populatie en verwijzing, verdient validatie met een volledig onafhankelijke dataset de voorkeur.

MRI biomarkers

Er is tot op heden vooral onderzoek gedaan naar diagnostische biomarkers, waarbij modellen zijn gemaakt om patiënten van gezonde mensen te onderscheiden. De klinische relevantie van dergelijke biomarkers is echter beperkt, omdat het weinig relevant is óf iemand een psychiatrische stoornis heeft, maar het vooral interessant is welke psychiatrische stoornis een patiënt heeft. Er zijn echter nog maar weinig studies naar differentiaaldiagnostiek die aan de hand van biomarkers een diagnose kunnen stellen, bijvoorbeeld voor het onderscheid tussen unipolaire en bipolaire depressie.⁶ De subjectiviteit van de klinische diagnostiek vormt echter een principiële probleem voor het maken van objectieve biomarkers. Het handboek voor de classificatie van psychische stoornissen (DSM-5) omvat

meer dan 250 afzonderlijke aandoeningen. Het is daarom niet verwonderlijk dat diagnostiek door twee verschillende klinici voor de meeste stoornissen niet goed overeenkomt,⁷ waardoor er geen 'gouden standaard' beschikbaar is voor het maken van biomarkers.

Prognostische biomarkers zijn daarom mogelijk relevanter voor de psychiatrie. Een veel onderzocht probleem is het voorspellen of patiënten met lichte psychotische symptomen uiteindelijk een psychotische stoornis zullen ontwikkelen. Wanneer het mogelijk is om patiënten met een hoog risico te identificeren, wordt het ook mogelijk om vroeg te interveniëren en een psychotische stoornis en de daarmee gepaarde functionele beperking te voorkomen.⁸ Een recente meta-analyse van zowel MRI als klinische data laat zien dat machine learning modellen redelijk goede prognoses voor individuele patiënten kunnen maken. Zo kunnen de modellen voorspellen of patiënten in de komende 1 tot 8 jaar een volledige psychotische stoornis ontwikkelen.⁹ De modellen hebben een gemiddelde sensitiviteit van 71% en specificiteit van 75%, wat betekent dat 71% van de patiënten met een latere stoornis en 75% van de patiënten zonder een latere stoornis correct worden geïdentificeerd. Dit lijkt al beter bruikbaar dan klinische vragenlijsten die weliswaar bijna alle psychotische patiënten kunnen identificeren (96% sensitiviteit), maar daarbij geen onderscheid kunnen maken tussen patiënten die later wel of geen stoornis zullen krijgen (47% specificiteit).¹⁰ Door het relatief kleine aantal patiënten dat uiteindelijk psychotisch blijft, worden daarmee te veel patiënten onterecht geïdentificeerd (fout positieven).

Andere veelbelovende toepassingen zijn voorspellende biomarkers die de uitkomst van behandelingen voorspellen. Doorgaans reageert ongeveer 50% van de patiënten op een behandeling. Dit effect is overigens vergelijkbaar met dat voor medicamenteuze behandeling van verschillende somatische aandoeningen.¹¹ Omdat het op dit moment nog niet te voorspellen is welke patiënt op welke behandeling zal reageren, wordt in de klinische praktijk een 'stepped-care' model gebruikt waarbij door 'trial-and-error' naar verbetering wordt gezocht. Na het uitproberen van vier verschillende behandelingen herstelt ongeveer twee derde van de patiënten met een depressie.¹² Het zou daarom behulpzaam zijn om elke patiënt meteen de juiste behandeling te kunnen bieden. Wij hebben onlangs een meta-analyse van machine learning studies uitgevoerd, waarbij de stemmingsverbetering middels MRI scans werd voorspeld. De gemiddelde sensitiviteit was 77% en de gemiddelde specificiteit 79%, met weinig verschil in nauwkeurigheid voor de verschillende typen behandelingen.¹³ Dit suggereert dat machine learning modellen goede voorspellingen kunnen maken voor individuele patiënten, wat de mogelijkheid creëert om MRI te gaan gebruiken bij de behandelindicatie voor patiënten.

Kanttekeningen

Een belangrijke kanttekening hierbij is echter dat bovenstaande meta-analyses op kleine steekproeven zijn gebaseerd. Machine learning modellen worden in principe altijd nauwkeuriger wanneer ze worden getraind op meer data, maar vaak lijkt het tegenovergestelde te gebeuren.¹⁴ Dit komt mogelijk doordat grotere studies ruimere inclusiecriteria hanteren en dat multicenter studies verschillende scanners gebruiken voor de dataverzameling. Dergelijke factoren vergrootten de verscheidenheid van een steekproef en van de data, waardoor machine learning modellen minder nauwkeurig worden.¹⁵ Daarnaast heeft elke studie een andere biomarker gebruikt. Het is daarom nog onbekend of een biomarker van ziekenhuis A ook bruikbaar is in ziekenhuis B. Dat lijkt in principe wel mogelijk, maar vooral wanneer de patiëntgroepen vergelijkbaar zijn.¹⁶ Toch zijn er ook meerdere

voorbeelden van nauwkeurige diagnostische biomarkers die niet goed generaliseren naar andere ziekenhuizen, zelfs wanneer ze gebaseerd zijn op grote steekproeven.¹⁷ Dat is op zich niet heel uitzonderlijk, aangezien ook verschillende referentiewaarden voor bloedonderzoek laboratoriumafhankelijk zijn (<https://www.nvkc.nl/algemeen-overzicht-referentiewaarden>). Het is dus altijd een optie om per ziekenhuis een andere biomarker te maken. Machine learning biedt ook de mogelijkheid om modellen per ziekenhuis aan te passen waarvoor slechts weinig nieuwe data nodig is ('transfer learning'), en om modellen te blijven optimaliseren aan de hand van nieuwe patiëntgegevens. Een dergelijke toepassing kan daardoor nauwkeuriger worden dan een standaard biomarker, maar kan daardoor niet 'out-of-the-box' worden gebruikt.

Een MRI scan kan vrijwel alleen in ziekenhuizen worden gemaakt en kost anno 2020 een kleine 300 euro. Klinische gegevens en vragenlijsten zouden mogelijk een eenvoudiger en goedkoper alternatief vormen. Echter, een meta-analyse suggereert dat de voorspellende waarde van neuroimaging data (MRI en elektro-encefalografie (EEG)) aanzienlijk hoger is (85%) dan die van klinische data (69%). Ook genetische data hebben geen hoge voorspellende waarde (68%).¹⁸ Hierbij moet echter wel worden opgemerkt dat de steekproefgroottes van deze typen onderzoek niet vergelijkbaar waren. Terwijl de neuroimaging studies slechts enkele tientallen deelnemers hadden, omvatten de studies met klinische data wel honderden tot duizenden deelnemers. Hierdoor mogen we aannemen dat de werkelijke voorspellende waarde van klinische variabelen niet veel van de bovengenoemde 69% zal afwijken, maar die voor neuroimaging technieken zou 10% hoger of lager kunnen uitvallen.¹⁴ Mogelijk biedt EEG een goed bruikbaar alternatief. De apparatuur is goedkoper en kan ook eenvoudig in GGZ instellingen worden geplaatst. Hoewel een eerdere meta-analyse nog een te lage gemiddelde sensitiviteit (72%) en specificiteit (68%) rapporteerde voor elke vorm van voorspellende EEG biomarkers voor depressie,¹⁹ lijken machine learning modellen betere voorspellingen te kunnen maken. Zo vond een relatief grote studie een correlatie van 0.6 tussen de voorspelde en daadwerkelijke stemmingsverbetering (Wu 2020),²⁰ wat vergelijkbaar is met een nauwkeurigheid van 80% voor dichotome uitkomsten.

Conclusie

MRI heeft de belofte voor klinische toepassing in psychiatrie tot op heden nog niet waargemaakt. MRI heeft de afgelopen decennia wel bijgedragen aan een beter beeld van de structurele en functionele hersenafwijkingen van patiënten met psychiatrische aandoeningen, maar de verschillen ten opzichte van gezonde mensen blijken bijzonder klein. De recente toepassing van machine learning analyses lijkt echter veelbelovend. Er is nog weinig bekend over diagnostische biomarkers die onderscheid kunnen maken tussen verschillende stoornissen, maar prognostische biomarkers voor psychose⁹ en voorspellende biomarkers voor depressie¹³ benaderen de voorspellende waarde van 80% wat de American Psychiatric Association bruikbaar acht voor klinische toepassing.³ Die 80% is echter slechts een richtlijn omdat de daadwerkelijke bijdrage moet blijken uit een kosten-batenanalyse. Gelet op het gemiddelde behandelingsucces van 50% op dit moment¹¹ zal deze kosten-batenanalyse waarschijnlijk ook bij een lagere voorspellende waarde positief uitvallen, aangezien de behandelingseffectiviteit toeneemt met de nauwkeurigheid van de biomarker. De belofte moet echter nog wel worden waargemaakt in grotere en representatieve steekproeven waarbij meerdere behandelopties kunnen worden afgewogen. En om toepassing ook in andere ziekenhuizen mogelijk te maken, dient de dataverzameling zoveel als mogelijk op voorhand te worden afgestemd. Dit vergt een investering in een representatieve nationale biobank om robuuste machine learning modellen te

maken. Die kunnen vervolgens de basis vormen van clinical decision support systems die klinici informatie verschaffen over de verwachte prognose of respons op verschillende behandelingen. Daarmee wordt evidence-based zorg op maat mogelijk, met sneller herstel voor de patiënt en lagere kosten voor de maatschappij als gevolg.

Referenties

1. Rogers W, Thulasi Seetha S, Refaee TAG, Lieverse RIY, Granzier RWY, Ibrahim A, et al. Radiomics: from qualitative to quantitative imaging. *Br J Radiol.* 2020;93:20190948.
2. Thompson PM, Jahanshad N, Ching CRK, Salminen LE, Thomopoulos SI, Bright J, et al. ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl Psychiatry.* 2020;10:100.
3. First M, Botteron K, Carter C, Castellanos FX, Dickstein DP, Drevets W, et al. Consensus report of the APA work group on neuroimaging markers of psychiatric disorders: American Psychiatric Association; 2012.
4. McTeague LM, Huemer J, Carreon DM, Jiang Y, Eickhoff SB, Etkin A. Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. *Am J Psychiatry.* 2017;174:676-85.
5. McTeague LM, Rosenberg BM, Lopez JW, Carreon DM, Huemer J, Jiang Y, et al. Identification of common neural circuit disruptions in emotional processing across psychiatric disorders. *Am J Psychiatry.* 2020;177:411-21.
6. Han KM, De Berardis D, Fornaro M, Kim YK. Differentiating between bipolar and unipolar depression in functional and structural MRI studies. *Prog Neuropsychopharmacol Biol Psychiatry.* 2019;91:20-7.
7. Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, et al. DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. *Am J Psychiatry.* 2013;170:59-70.
8. McGorry PD, Hartmann JA, Spooner R, Nelson B. Beyond the "at risk mental state" concept: transitioning to transdiagnostic psychiatry. *World Psychiatry.* 2018;17:133-42.
9. Sanfelici R, Dwyer DB, Antonucci LA, Koutsouleris N. Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: A meta-analytic view on the state of the art. *Biol Psychiatry.* 2020;88:349-60.
10. Fusar-Poli P, Cappucciati M, Rutigliano G, Schultze-Lutter F, Bonoldi I, Borgwardt S, et al. At risk or not at risk? A meta-analysis of the prognostic accuracy of psychometric interviews for psychosis prediction. *World Psychiatry.* 2015;14:322-32.
11. Leucht S, Hierl S, Kissling W, Dold M, Davis JM. Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. *Br J Psychiatry.* 2012;200:97-106.
12. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry.* 2006;163:1905-17.
13. Cohen SE, Zantvoord JB, Wezenberg BN, Bockting CLH, van Wingen GA. Magnetic resonance imaging for individual prediction of treatment response in major depressive disorder: a systematic review and meta-analysis. *MedRxiv preprint.* 2020:doi:10.1101/2020.06.27.20141465.
14. Varoquaux G. Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage.* 2018;180:68-77.
15. Schnack HG, Kahn RS. Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters. *Front Psychiatry.* 2016;7:50.
16. Jiang R, Abbott CC, Jiang T, Du Y, Espinoza R, Narr KL, et al. SMRI Biomarkers Predict Electroconvulsive Treatment Outcomes: Accuracy with Independent Data Sets. *Neuropsychopharmacology.* 2018;43:1078-87.
17. Lei D, Pinaya WHL, van Amelsvoort T, Marcelis M, Donohoe G, Mothersill DO, et al. Detecting schizophrenia at the level of the individual: relative diagnostic value of whole-brain images, connectome-wide functional connectivity and graph-based metrics. *Psychol Med.* 2019:1-10.

18. Lee Y, Ragguett RM, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, et al. Corrigendum to "Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review." *J Affect Disord.* 241 (2018) 519-532. *J Affect Disord.* 2020;274:1211-5.
19. Widge AS, Bilge MT, Montana R, Chang W, Rodriguez CI, Deckersbach T, et al. Electroencephalographic Biomarkers for Treatment Response Prediction in Major Depressive Illness: A Meta-Analysis. *Am J Psychiatry.* 2019;176:44-56.
20. Wu W, Zhang Y, Jiang J, Lucas MV, Fonzo GA, Rolle CE, et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nat Biotechnol.* 2020;38:439-47.