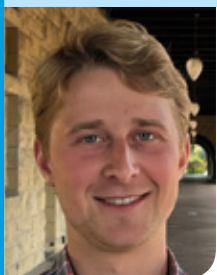


VAN RUIS NAAR REALITEIT

Fictieve data met echte impact



Lennart Koetzier



Martin Willeminck

Om betrouwbare AI-modellen te trainen, zijn grote, diverse en representatieve datasets nodig. Een veelbelovende oplossing is het gebruik van synthetische data: door AI gegenereerde beelden die echte medische beelddata nabootsen, maar geen herleidbare patiënt-informatie bevatten.

Aan de basis van veel medische AI-tools liggen zogeheten generatieve *foundation models*. Deze basismodellen maken het mogelijk om nieuwe content te genereren op basis van bestaande data. Deze modellen vragen echter om miljoenen trainingsbeelden. In de medische praktijk is dat een uitda-

originele beeld. Dit getrainde model wordt vervolgens ingezet om vanuit een ruisvector realistische, synthetische beelden te genereren (zie figuur 1). Diffusiemodellen bieden de hoogste beeldkwaliteit en veel controle over het gegenereerde resultaat, al vergen ze meer rekenkracht dan de eerdergenoemde modellen.

ta. Zo demonstreerde Pan et al. dat een covid-19-classicator, getraind op diffusiemodelgegenereerde synthetische thoraxfoto's, even goed presteerde als bij echte data en dat extra synthetische beelden de accuratesse verhoogden van 89 naar 93%.³

Een tweede toepassing is modaliteitstranslatie. Intramodaliteitstranslatie blijft binnen dezelfde modaliteit, maar wijzigt acquisitie- of reconstructie-instellingen: zo transformeerde een GAN-synthetische MRI-sequenties op basis van echte MRI-sequenties, waarbij de synthetische sequenties visueel en kwantitatief nauwelijks van de echte beelden verschilden (zie figuur 2).⁴ Intermodale translatie zet beelden om tussen modaliteiten, bijvoorbeeld van MRI naar CT (of omgekeerd). Hierdoor wordt stralingsdosis vermeden of is met 1 opname zowel bot- als weke-deleninformatie te verkrijgen. Bij contrastsynthese transformeert een generatief model een CT- of MRI-scan met non-contrast of lage contrast naar syn-

‘Met generatieve modellen zijn realistische beelden met zeldzame of subtiele afwijkingen te maken. Dit verrijkt de opleiding van radiologen’

ging door privacywetgeving, gefragmenteerde datasystemen en een gebrek aan gestandaardiseerde infrastructuur.¹ Synthetische data bieden voordelen, maar ook risico's.²

Synthetische beelden

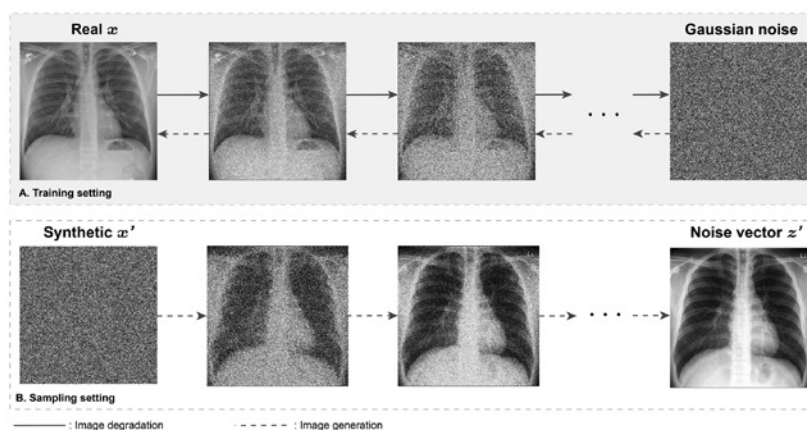
Synthetische beelden worden traditioneel gegenereerd met *generative adversarial networks* (GAN's) en *variational auto-encoders* (VAE's). Deze modellen leren op basis van bestaande beelddata om nieuwe beelden te creëren met vergelijkbare anatomische, pathologische en visuele kenmerken. Belangrijk is dat deze beelden statistisch representatief zijn, maar niet te herleiden zijn tot individuele patiënten.

Diffusiemodellen

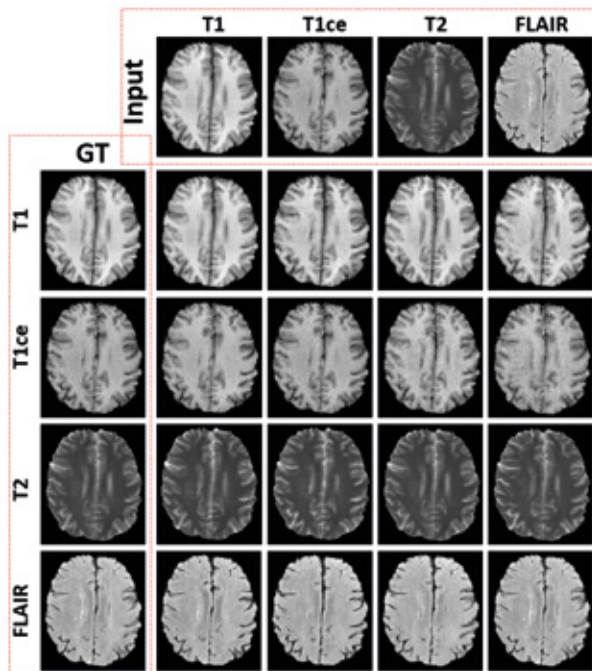
Een meer recente ontwikkeling zijn diffusiemodellen, die in trainingsfase extreme ruis toevoegen aan echte beelden, tot slechts ruis overblijft. Deze modellen leren ruis daarna iteratief te verwijderen om weer te komen tot het

Medische beeldvorming

De eerste toepassing is datasetverrijking. AI-modellen presteren vaak beter wanneer zeldzame klassen, zoals bepaalde tumortypen of ondervertegenwoordigde demografische groepen, synthetisch worden toegevoegd aan de trainingsda-



Figuur 1. Een diffusiemodel leert door echte beelden stapsgewijs met ruis te degraderen en vervolgens de ruis te onderdrukken, zodat het bij sampling vanuit willekeurige ruis dezelfde keten terugloopt en volledig nieuwe, realistische beelden kan synthetiseren.



Figuur 2. De bovenste rij toont de inputsequenties, de eerste kolom de ground-truthbeelden; overige cellen zijn synthetische intramodale translaties.

thetische contrastbeelden en kan het zo de diagnostische waarde verhogen of kosten verminderen.⁵

Onderwijs en training

Een derde toepassing is onderwijs en training. Met generatieve modellen zijn realistische beelden met zeldzame of subtiele afwijkingen te maken. Dit verrijkt de opleiding van radiologen. Gecombineerd met explainable AI-technieken kunnen gebruikers visualiseren welke beeldkenmerken een AI-detectiemodel gebruikt voor een voorspelling. Door bijvoorbeeld de voorspelling voor cardiomegalie doelbewust te versterken en de resulterende gradiënt terug te propageren, vergroot het model het hart in het beeld. Zo worden correlaties zichtbaar gemaakt en begrijpen gebruikers welke beeldkenmerken werkelijk bepalend zijn, en groeit het vertrouwen in AI-ondersteunde besluitvorming.

Privacy en data-uitwisseling

Synthetische beelden kunnen mogelijk privacybarrières overbruggen in het delen van data, doordat ze geen herleidbare informatie bevatten en dus makkelijker deelbaar zijn voor onderzoek en ontwikkeling van AI. Studies tonen echter dat sommige modellen onbedoeld herkenbare patronen repliceren. Dat maakt re-identificatie mogelijk. Strikte evaluatie en validatie blijven daarom essentieel, vooral buiten de instelling waar de beelden zijn gegenereerd.

Beperkingen

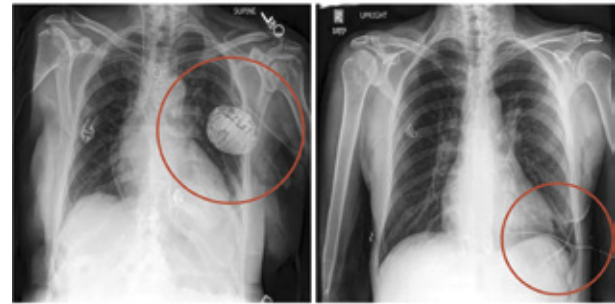
Hoewel de potentie groot is, zijn er ook uitdagingen:

- **Realisme en diversiteit:** synthetische beelden moeten geloofwaardig en gevarieerd zijn om te voorkomen dat AI-modellen verkeerde correlaties leren. Hiervoor moeten globale en lokale pathologische kenmerken realistisch in synthetische beelden voorkomen zonder dat onrealistische kenmerken worden toegevoegd (zie figuur 3).
- **Bias en representativiteit:** synthetische data worden gegenereerd op basis van bestaande datasets, die mogelijk al gebiast zijn. Dit kan leiden tot versterking van demografische ongelijkheden.
- **Regelgeving:** huidige wetgeving biedt onvoldoende houvast voor veilig gebruik, waardoor 1) herziening van bestaande privacywetten met duidelijke definities, 2) oprichting van een centrale toezichthouder en 3) transparantie bij ontwikkeling en toepassing dringend noodzakelijk zijn.

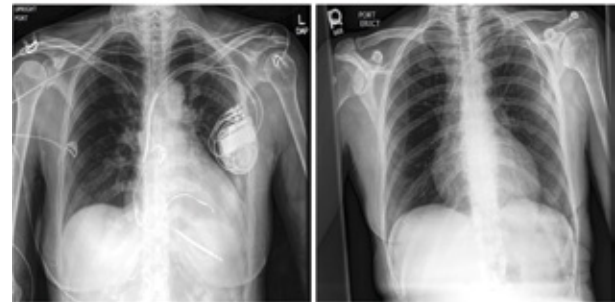
Nauwe samenwerking

Samenvattend: synthetische data kunnen het ontwikkelen van AI in de radiologie versnellen, met name op het gebied van dataverrijking, onderwijs, privacybescherming en workflowoptimalisatie. Echter vergt verantwoorde inzet van synthetische beelden gestandaardiseerde evaluatiemethoden, transparantie in ontwikkeling en gebruik van de modellen, en wetgeving die patiëntveiligheid en

a. Synthetic epD-GAN Chest Images



b. Real Nearest Neighbours



Figuur 3. (A) synthetische beelden met rode cirkels rond onrealistische halucinaties; (B) de dichtstbijzijnde echte trainingsbeelden ter vergelijking.

ethiek borgt. Dit vereist nauwe samenwerking tussen AI-ontwikkelaars, klinici en regelgevers. Radiologen spelen een sleutelrol als eindgebruikers én beoordelaars. Met hun betrokkenheid en een open, transparant proces kan synthetische beeldvorming uitgroeien tot een betrouwbare, breed toepasbare innovatie in de dagelijkse klinische praktijk. ■

Lennart Koetzier

promovendus in het UMC Utrecht

Martin Willeminck

medeoprichter en chief scientific officer bij Segmed

Referenties

1. Willeminck MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology*. 2020 Apr;295(1):4-15.
2. Koetzier LR, Wu J, Mastrodicasa D, et al. Generating synthetic data for medical imaging. *Radiology*. 2024 Sep;312(3):e232471.
3. Pan S, Wang T, Qiu RLJ, et al. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Phys Med Biol*. 2023 May 5;68(10):105004.
4. Liu X, Xing F, Fakhri GE, et al. A unified conditional disentanglement framework for multimodal brain mr image translation. *Proc IEEE Int Symp Biomed Imaging*. 2021 Apr;2021:10.1109/isbi48211.2021.9433897.
5. Azarfar G, Ko SB, Adams SJ, et al. Applications of deep learning to reduce the need for iodinated contrast media for CT imaging: a systematic review. *Int J Comput Assist Radiol Surg*. 2023 Oct;18(10):1903-14.